Glycan Scavenging at the Cell Surface: The Bacteroidetes Sus-like paradigm

by

Matthew Harrison Foley

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in the University of Michigan
2018

Doctoral Committee:

Assistant Professor Nicole M. Koropatkin, Chair
Research Assistant Professor Chris J. Alteri
Associate Professor Julie S. Biteen
Professor Vernon B. Carruthers
Associate Professor Eric C. Martens

Matthew Foley

mhfoley@umich.edu

ORCID iD: 0000-0002-3116-2040

Acknowledgements

I would first like to thank my thesis mentor Dr. Nicole Koropatkin for her unbounded support and guidance over the past four years. Thank you for demonstrating everyday that enthusiasm, personality, and kindness can positively influence your science and research. Thank you for trusting and investing in me from the beginning of my training. Thank you for allowing me to pursue unique challenges and training opportunities. Thank you for genuinely caring about my well-being and mental health. Thank your for being a positive role model in my life. Most of all, thank you for being a friend.

I would like to thank my thesis committee members Dr. Eric Martens, Dr. Vernon Carruthers, Dr. Chris Alteri, and Dr. Julie Biteen for their insights, guidance, and collaboration throughout my training. I would especially like to thank Dr. Eric Martens for his support, help, feedback, and interest in my work over the years.

I would like to thank all current and past members of the Koropatkin and Martens labs for their help and advice, in particular Constance Bahr and Dr. Darrell Cockburn for their friendship and for making the lab a fun place to work and learn. My collaborators Dr. Hannah Tuson, Dr. Julie Biteen, Dr. Guillaume Déjean, Kazune Tamura, and Dr. Harry Brumer have all greatly contributed to my thesis work. I would also like to thank the University of Michigan Department of Microbiology and Immunology for providing a wonderful and collaborative environment for me to train in.

I would like to thank the Cellular Biotechnology Training Program for partially funding the work described in this thesis, and also for the unique leadership and professional experiences they have given me. I am additionally grateful for the mentorship and experience received by Dr. Michael Ford at MS Bioworks who offered me new perspectives on science and industry.

Finally, I would like to thank my parents, family, and friends for their support inside and outside of the lab. Courtney Luterbach, Zack Abbott, Hayley Warsinske, Travis Kochan, Danelle Weakland, Kyle Luterbach, David Collins, Joseph Schuman, Christine Biddlecombe, Jordan Johnson, Zach Barber, Allie Dolin, Emma Decotis, Terri Kurtz, and Matt Hudecek all have contributed to my success and achievements. I am especially appreciative for the help and love received by Courtney Luterbach.

Table of Contents

List of Tables

List of Figures

List of Movies

Abstract

Resident bacteria in the densely populated human intestinal tract must efficiently compete for carbohydrate nutrition. The Bacteroidetes, a dominant bacterial phylum in the mammalian gut, encode a plethora of discrete polysaccharide utilization loci (PULs) that are selectively activated to facilitate glycan capture at the cell surface. The most well-studied PUL-encoded glycan uptake system is the starch utilization system (Sus) of *Bacteroides thetaiotaomicron.* The Sus includes the requisite proteins for binding and degrading starch at the surface of the cell preceding oligosaccharide transport across the outer membrane for further depolymerization to glucose in the periplasm. In the Sus of *B. thetaiotaomicron*, the well-characterized outer membrane proteins SusDEF and the α-amylase SusG each have unique structural features that allow them to interact with starch. The work presented here will address several important questions regarding how SusCDEFG and their homologs in other Sus-like systems work in relation to each other and with their cognate glycans to further develop what is known about the Sus-like paradigm of nutrient acquisition that is exclusive to the Gram-negative Bacteroidetes.

Despite the apparent redundancy in starch-binding sites among these proteins, each has a distinct role during starch catabolism. In this thesis, new roles for the starch-binding protein SusE and SusF are described that further our understanding of maltooligosaccharide transport in the Sus. Sus outer membrane proteins dynamically interact and cooperate in the membrane. A new model is proposed for the formation of a Sus complex that assembles around the stationary starch-binding proteins SusE and

SusF. Making comparisons across diverse PUL allow for the understanding of the conserved and divergent features of Sus-like systems. We demonstrate here that the Xyloglucan Utilization Locus of *Bacteroides ovatus* contains a hierarchy of xyloglucan-binding proteins that facilitate polysaccharide uptake and a cell surface GH9 plays a previously unappreciated role in xyloglucan capture. Lastly, the *B. ovatus* Mixed-Linkage Glucan Utilization Locus cell surface glycan-binding proteins are shown to display different contributions to mixed-linkage glucan capture and oligosaccharide uptake.

These studies of polysaccharide-targeting systems provide mechanistic insights into the underpinnings of PUL-encoded glycan capture, and more broadly, how a fundamental phylum of the gut microbiome acquires carbohydrate nutrition

Chapter I: Introduction

Notes

Pages 1 – 29 and 33 – 35 of this chapter were reprinted and modified with permission from Foley, M.F., Cockburn, D.W., Koropatkin, N.M. The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cellular and Molecular Life Sciences*. **73**, 2603-2617 (2016).

The Sus operon: a model system for starch uptake by the human gut Bacteroidetes

The consortium of bacteria that inhabit the mammalian gastrointestinal tract have a profound influence on host development and health [1-4]. A notable function of these microbes is the digestion and fermentation of both endogenous (i.e., host-derived) and dietary carbohydrates into short chain fatty acids that offer a physiological benefit to the host [2, 5]. The bacteria that have adapted to persist and thrive in this densely populated ecosystem have evolved efficient strategies to harvest glycans, and it is the competition and synergy among species for their preferred glycans that drives the diet-dependent changes observed in the gut community structure [6-9].

Starch is produced by plants as an energy storage compound and is the dominant carbohydrate component of most Western style diets. It is produced by the plants as granules made up of two polymers of glucose, the linear α(1,4)-linked amylose and the branched amylopectin with an α(1,4)-linked backbone and α(1,6) branch points

[10]. A recent analysis of the glycolytic potential encoded within the genomes of gut bacteria using the Carbohydrate-Active enZYme (CAZy) database (www.CAZy.org) reflected that genes encoding starch-processing enzymes from the glycoside hydrolase family 13 (GH13) are among the most represented in the microbiota [11]. This broad potential for starch utilization is distributed among the Bacteroidetes, Firmicutes and Actinobacteria, the three most abundant phyla of gut bacteria. Humans are able to efficiently process most starch and it is only the resistant starch fraction of a more crystalline nature which survives transit to the large intestine where it is exposed to the gut microbiota [12]. The ability of these organisms to each thrive on specific forms of starch (e.g., complex resistant starch granules, soluble maltooligosaccharides, and amylopectin) is dependent upon both the specific activity of their GH13 enzymes and the types of glycan-uptake systems that work in concert with these enzymes. The enzymes used by the microbiota for starch degradation fall into three broad classes: amylases/neopullulanases that act upon the α(1,4)-linkages, pullulanases that act upon the α(1,6)-linkages, and α-glucosidases that act upon both types of linkages releasing glucose, typically from oligosaccharides [13]. To study these enzymes a variety of model substrates are used. This includes isolated amylose, amylopectin, maltooligosaccharides, pullulan (an α(1,6)-linked maltotriose polysaccharide), and cyclodextrins (circularized oligosaccharides that mimic the helical shape of amylose and amylopectin). Descriptions of these substrates are summarized in Table 1.1.

Table 1.1. Characteristics of starch-related carbohydrates

Carbohydrate	Description
Amylose	A starch polymer comprised of α(1,4)-linked glucose. The a(1,4) glycosidic linkage creates a helical conformation in solution and in the starch granule. Amylose helices can pack together creating insoluble crystalline regions within a starch granule.
Amylopectin	The branched starch molecule differentiates itself from amylose by containing α(1,6)-linkage branch points along the α(1,4)-linked glucose backbone. These branches prevent the tight packing of neighboring helices resulting in amorphous regions within the starch granule and enhanced solubility.
Maltooligosaccharides	Oligosaccharides of starch that are typically generated by amylolytic enzymes operating on the full-length polysaccharide. Purified oligosaccharides of known length allow for the more precise study of protein-carbohydrate binding and activity.
Pullulan	A linear starch-like polysaccharide containing repeating units of α(1,6)-linked maltotriose. The α(1,6)-linkages may mimic branch points in amylopectin and they are sometimes used to determine an enzyme's tolerance or activity towards those branch points.
Cyclodextrins	Cyclic oligosaccharides of a(1,4)-linked glucose that mimic the curvature of a starch helix. The extent of this curvature, and similarly the molecule's constrained geometry, decreases as the number of glucoses in the oligosaccharide increases. Most commonly used cyclodextrins include α-cyclodextrin and β-cyclodextrin that contain six and seven glucose residues, respectively, because of their similarity to the curvature of a starch helix.
Resistant Starch (RS)	Starch that is impervious to degradation by human dietary amylases due to inaccessibility, crystallinity, chemical modifications, or complex formation with lipids. RS becomes available to colonic microorganisms that are either equipped with the molecular machinery to degrade RS themselves or are available to crossfeed from RS-degrading organisms.

In addition to the diversity of enzymes employed, the strategies used to capture hydrolyzed starch in the gut are a function of the unique physiology of the respective microorganisms [14-17]. The Gram-positive Firmicutes and Actinobacteria take up monosaccharides and oligosaccharides via a variety of transport systems including ATP-binding cassette transporters, major facilitator superfamily, and phosphotransferase systems [18, 19]. Many of these transporters are encoded within putative operons that include one or more extracellular GH13 enzymes to hydrolyze starch at the cell surface [14, 17, 20]. In contrast, the genomes of most Bacteroidetes, the dominant Gram-negative phylum in the mammalian gut, have far fewer of these classically studied carbohydrate-uptake systems [21]. Rather the Bacteroidetes package their glycolytic potential within discrete gene clusters termed polysaccharide utilization loci (PUL) that encode glycoside hydrolases, glycan-binding proteins and a TonB-dependent transporter [15]. These PUL-encoded proteins work together in the outer membrane to capture and transport glycans, including starch.

Abigail Salyers' seminal work on the carbohydrate-degrading phenotypes of human gut *Bacteroides* species laid the foundation for the modern study of this clade of bacteria. One of Salyers' first studies revealed that 22 isolates of *Bacteroides thetaiotaomicron* shared the ability to grow on amylose and amylopectin [22]. Further investigation of starch utilization by *B. thetaiotaomicron* VPI-5482 revealed that starch-binding to the cell surface was a pre-requisite to growth on starch, and was mediated by several proteins in the outer membrane [23]. Through their efforts to determine the molecular basis of starch utilization, the Salyers lab identified the first PUL, an eight-gene cluster in *B. thetaiotaomicron* that encodes all of the required proteins for starch

adherence to the cell, as well as a surface amylase and TonB-dependent transporter to coordinate starch hydrolysis with maltooligosaccharide import into the cell (**Figure 1.1**). This gene cluster was named the starch utilization system (Sus) for its apparent function, and is composed of *susRABCDEFG* [24, 25]. SusR is an inner membrane-spanning sensor/regulator protein that recognizes maltose, a disaccharide of glucose, in the periplasm and triggers the rapid upregulation of the *sus* genes [25]. The outer membrane lipoproteins SusDEF facilitate the binding of starch to the cell surface, and bound starch is then hydrolyzed by the α-amylase SusG [26-28]. The resulting maltooligosaccharides are shuttled into the periplasm via SusC, a TonB-dependent transporter [29], and further depolymerized by the neopullulanase SusA and α-glucosidase SusB [30, 31].

A decade or more after Salyers' discovery, bacterial genome sequencing revealed that all gut Bacteroidetes possess PULs, and each PUL confers the ability to grow on a different glycan [15, 32]. All PULs encode homologues of the proteins SusCD, glycan-binding lipoproteins akin to SusEF, and a cadre of glycoside hydrolases for the utilization of a distinct glycan [15]. Based upon this commonality, all PUL-encoded proteins are believed to form a "Sus-like" system of proteins in the outer membrane that work together to target a specific glycan. PULs encoding Sus-like systems have been identified for the uptake of diverse substrates such as xyloglucan, arabinoxylan, α-mannan, inulin, and porphyran, among others [33-39]. Organisms such as *B. thetaiotaomicron* and *Bacteroides ovatus* dedicate ~18% of their genomes to PULs, highlighting the importance of the PUL-encoded glycan uptake strategy to the

adaptation of these organisms to the gut [39]. The repertoire of different PULs encoded by an organism dictates the metabolic lifestyle of the bacterium in the gut [39, 40].

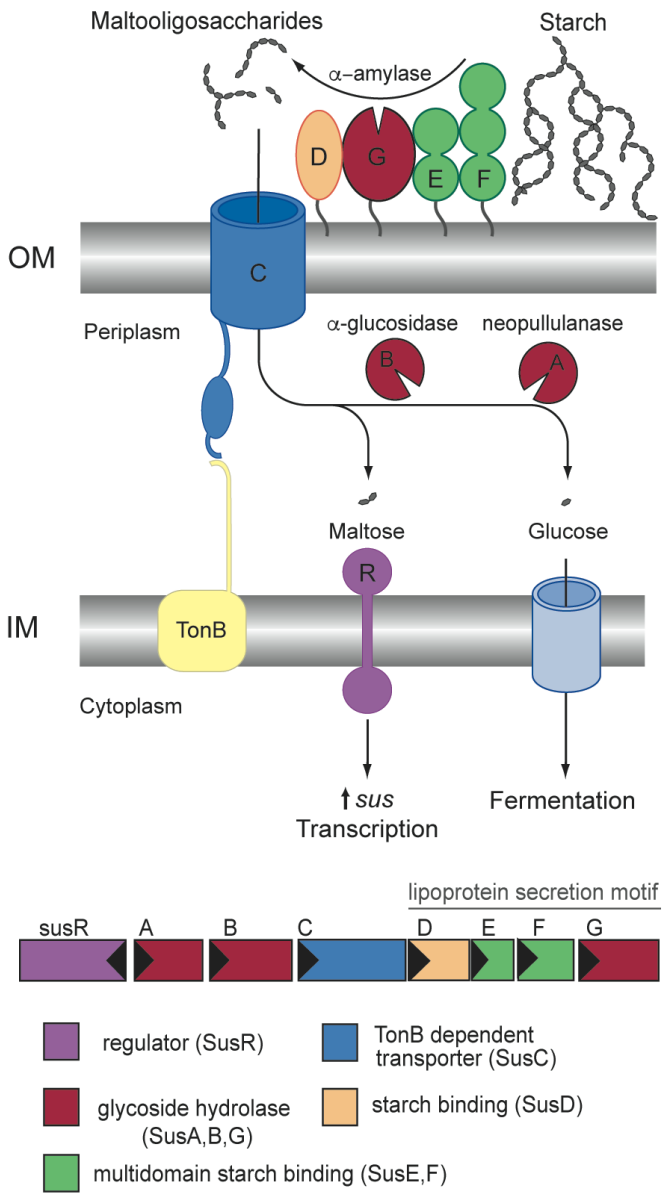

Figure 1.1. Overview of the starch utilization system (Sus) in B. thetaiotaomicron

The *sus* locus is transcribed from two divergent promoters. Transcription of *susR* occurs independently from the rest of the locus, which allows the inner membrane-spanning protein SusR to sense the disaccharide inducer, maltose, in the periplasm and subsequently drive the transcription of *susABCDEFG*. Starch is bound to the surface of the cell by the starch-binding outer membrane lipoproteins SusDEF. Subsequent hydrolysis by a similarly membrane-tethered α-amylase, SusG, generates oligosaccharides small enough to transit through the TonB-dependent transporter. Once in the periplasm, SusA and SusB, a neopullulanase and α-glucosidase, respectively, process oligosaccharides into glucose. The monosaccharide is then shuttled into the cytoplasm by an unknown transporter. The stoichiometry and assembly of the Sus is unknown.

The Sus of *B. thetaiotaomicron* remains the best-studied PUL-encoded glycan uptake system to date, and is often the prototypical system by which the function of homologous PUL-encoded proteins are compared or inferred. Here we summarize the structural and functional work to date on the Sus proteins of *B. thetaiotaomicron* VPI-5482, with a particular focus on the outer membrane proteins SusDEFG. These are not only the most well characterized proteins in the system, but together they exemplify the molecular strategy that the Bacteroidetes utilize to sense and acquire carbohydrate nutrition. These studies have shaped a general model of the "Sus-like" paradigm that dominates glycan catabolism by the mammalian gut Bacteroidetes.

SusD: An α-helical carbohydrate-binding protein

Starch adherence to the cell surface is the first step in starch utilization by *B. thetaiotaomicron* [23]. Salyers and colleagues used a polar insertion at *susE* (Ω*susE*) to create *B. thetaiotaomicron* that expressed only SusC and SusD, and noted that this mutant bound radioactively labeled starch at ~70% the levels of wild-type cells. This strain grew normally on starch if complemented with SusG (Ω*susE::susG)* [27]. However, neither SusC nor SusD alone could drive starch adsorption, as *B. thetaiotaomicron* cells expressing only one of these two proteins displayed barely detectable levels of starch-binding [41]. Furthermore, while neither isolated SusC nor SusD protein bound to amylose resin, mixing of the proteins prior to incubation with amylose resin resulted in the retention of both proteins [26]. The adaptation of a TonB-dependent transporter, a family of proteins historically associated with iron uptake, for

starch utilization as well as the requirement of the co-receptor protein SusD marked two novel features of this system.

While SusD aids in starch-binding to the cell surface, it has no amino acid sequence similarity to known carbohydrate-binding modules (CBMs), and is notably larger (65 kDa) than any carbohydrate-binding protein. However, bacterial genome sequencing revealed the ubiquitous inclusion of homologs of *susD* as well as *susC* within all PULs of the gut Bacteroidetes, suggesting a conserved function for the encoded proteins in glycan uptake [38]. The crystal structure of SusD revealed an abundance of α-helices, with a single starch-binding site [42]. SusD is tethered to the outer membrane via a lipidated N-terminal cysteine preceded by a 16-residue flexible linker, effectively projecting the protein above the surface of the membrane like a balloon on a string. (As discussed in later sections, lipidation followed by a flexible linker is a conserved feature of SusEFG as well). The most definitive feature of SusD and its homologs is the conservation of four helix-turn-helix motifs known as tetratricopeptide repeats (TPR), that form a right-handed superhelix along one face of the protein [43, 44] (**Figure 1.2A**, in darker colors). TPR motifs are ubiquitous in nature and most commonly support protein complex formation by serving as a site for protein-protein interactions [44]. The TPR portion of SusD-like proteins is almost structurally invariant and serves as a scaffold for the more variable remainder of the protein that includes the ligand-binding site [43].

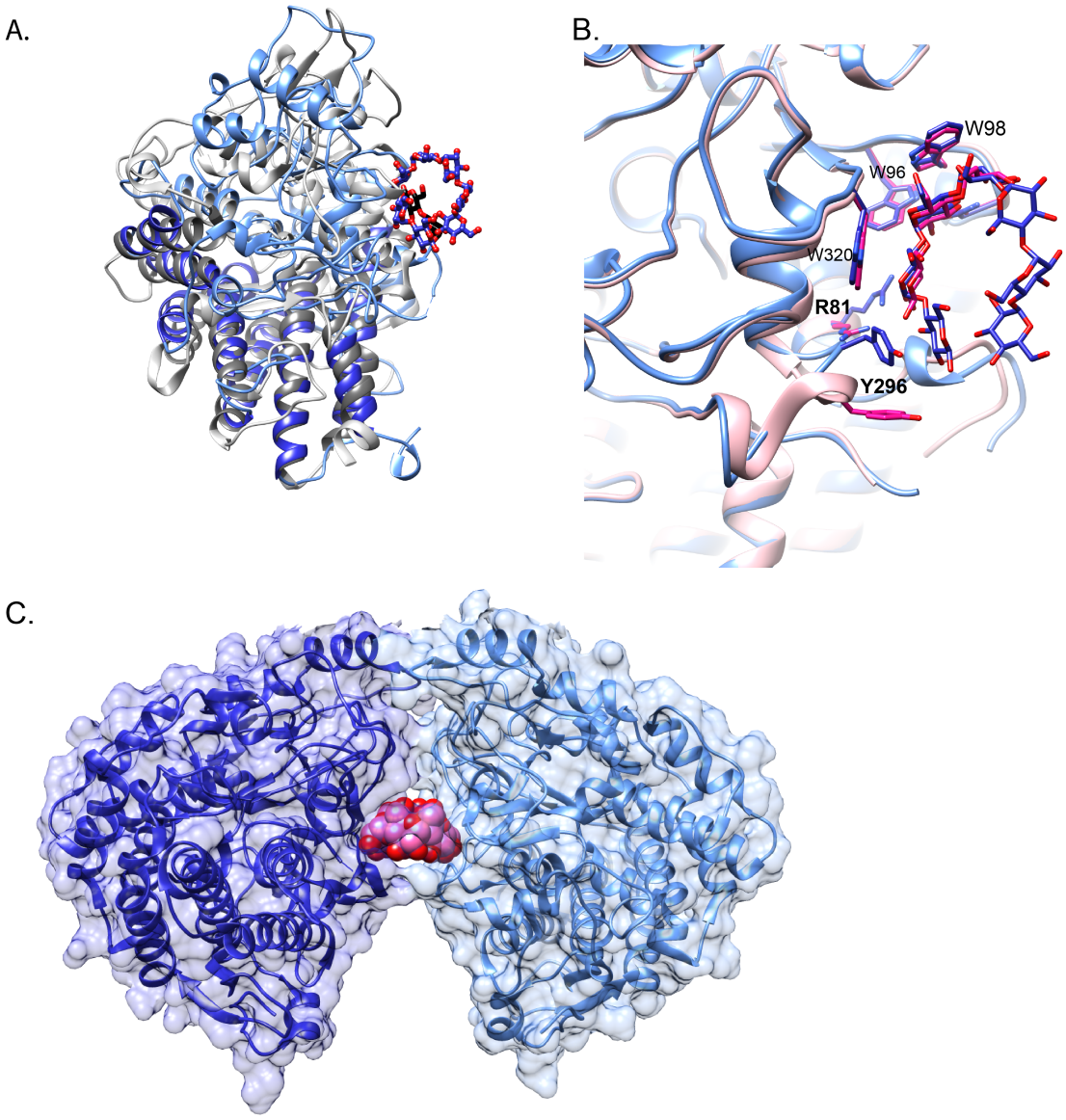

Figure 1.2. Molecular structure of SusD with maltooligosaccharides

A. Superposition of SusD (blue, PDB 3CK9) with bound maltoheptaose (blue sticks) and the SusD homolog BT1043 (grey, PDB 3EHN) that targets mucosal glycans with bound N-acetyllactosamine (black sticks). The conservation of the eight tetratricopeptide repeat helices is highlighted in darker colors for both proteins. The RMSD for these proteins is 2.8 Å over 324 aligned residues (12.6% sequence identity). B. Superposition of the structure of SusD with bound maltoheptaose (blue) and maltotriose (pink), highlighting the plasticity within the binding site. Residues that move upon binding of a longer α-glucan are in bold print. C. SusD crystallized with α-cyclodextrin revealed protein dimerization. The affinity of starch-binding to the cell surface may be enhanced by an avidity effect, whereby multiple SusD proteins cooperate to bind the polymer. The starch-binding site of SusD is a shallow pocket containing an arc of aromatic amino acids that complement the shape of an amylose helix [42]. The crystal structure of SusD with maltoheptaose reveals these residues as W96, W320 and Y296, with hydrogen bonding of the O2 and O3 hydroxyls of adjacent glucose residues via the side chains of R81

and N101 [42] (**Figure 1.2B**). These molecular determinants of starch recognition are shared with most starch-binding CBMs [45, 46], as well as the surface starch-binding sites of some GH13 enzymes [47, 48]. Thus, the glycan binding site of SusD may be an example of convergent evolution whereby proteins of distinct evolutionary lineage and hence structure evolve similar functions [49]. A unique feature of maltooligosaccharide recognition by SusD is the flexibility of two loops near the binding cleft, one of which includes Y296 that is part of the aromatic arc. The crystal structure of SusD with maltotriose suggests that glycan binding is initiated at W98 and W320, followed by a shift in these two flexible loops, one of which moves away from the binding site to allow Y296 to shift into position. This plasticity in the binding site of SusD may allow the flexible recognition of the α-glucan helix in the context of naturally occurring α(1,6)-branching.

The affinity of SusD for maltoheptaose (K_D ~ 1.0 mM) is much worse than most starch-binding CBMs that recognize maltoheptaose with low μM affinity [50-52]. Moreover, SusD displays negligible affinity for maltopentaose, and no detectable recognition of smaller sugars [42]. The binding affinity of SusD is akin to the surface starch-binding site of barley α-amylase [53], or that of the low-affinity starch-binding CBM45 family [54]. However, SusD binds β-cyclodextrin with a K_D ~ 0.15 mM, highlighting that this protein is more adapted to recognize a constrained helical structure such as a starch polymer over a flexible oligosaccharide [42]. At the cell surface, it is unknown how the interaction of SusD with SusC influences the cell's affinity for starch and maltooligosaccharides. The crystal structure of SusD with α-cyclodextrin displayed a glycan-induced dimerization, which hints at the potential of multiple SusDs to interact with a single ligand [42] (**Figure 1.2C**). Such an avidity effect could enhance the ability of SusD to facilitate starch-binding to the cell surface. Another possibility is that SusC increases the affinity of SusD for starch, either by inducing a conformational change in SusD that enhances binding, or by extending the protein-carbohydrate interaction in a complex between the two proteins [27].

Despite its relatively weak affinity for its ligand, SusD has a critical role in starch utilization: cells with an in-frame deletion of susD (*ΔsusD)* cannot grow on starch or maltooligosaccharides larger than maltopentaose, and display an intermediate growth phenotype on maltopentaose and maltotetraose [42]. The Sus proteins are dispensable for growth on maltotriose and maltose, which presumably enter the outer membrane via a non-specific porin [24]. More recent work following the discovery of SusE and SusF as

additional starch-binding proteins has revealed that the importance of SusD may extend beyond its ability to bind starch, as detailed in a later section.

SusG is a novel GH13 amylase required for starch utilization

Bacterial growth on polysaccharides including starch requires cell surface or extracellular glycoside hydrolases to break down the polymer into oligosaccharides that can be transported into the cell [55]. In *B. thetaiotaomicron*, SusG is the only cell surface amylase that is required for growth on starch [28]. Like SusD, SusG is tethered to the surface at an N-terminal cysteine followed by a 20 residue flexible linker before the first β-strand is formed [56]. While SusG displays the typical GH13 amylase family protein fold comprised of A, B, and C domains, a CBM58 is uniquely inserted within the B-domain sequence (**Figure 1.3A**, CBM in pink) [56]. The CBM58 is extended from the rest of the catalytic domain via two short β-strands and does not interact with the rest of the protein, creating an overall bilobed appearance. The unique placement of this CBM within the catalytic fold contrasts with the typical N or C-terminal placement of a starch-binding CBM within GH13 enzymes where the CBM can pack against the catalytic domain, sometimes shaping the active site or allowing dimerization [57-60].

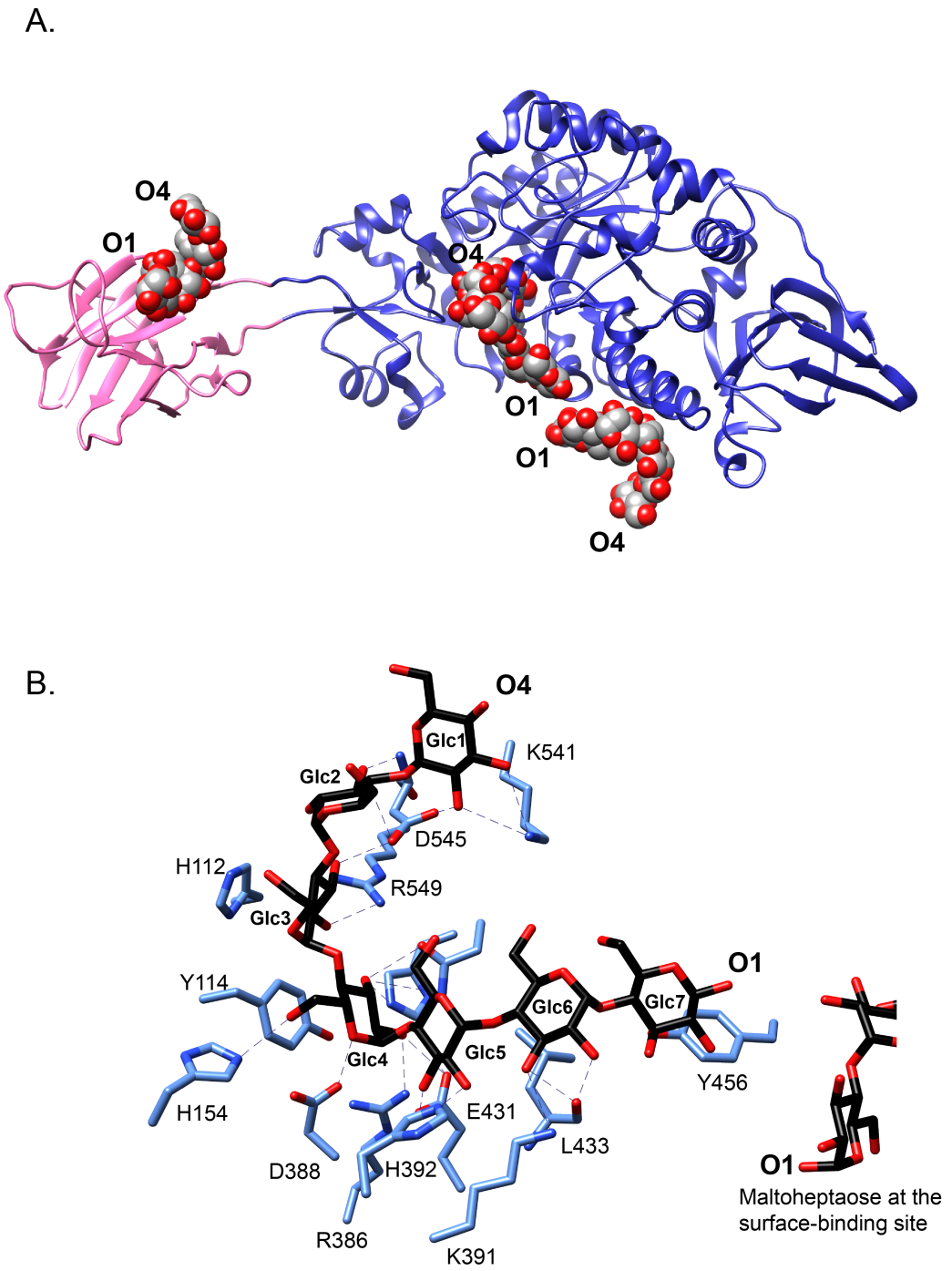

Figure 1.1. SusG is an amylase with a unique CBM insertion

A. Structure of the catalytically inactive mutant of SusG D498N (PDB 3K8L) with bound maltoheptaose. CBM58 (residues 216-335) is highlighted in pink, and maltooligosaccharides bound at CBM58, the active site, and the surface starch-binding site are depicted as spheres. The orientation of the oligosaccharide from the nonreducing end (O4) to reducing end (O1) is indicated. B. Close-up view of the active site in the catalytically inactive mutant of SusG D498N (PDB 3K8L) with bound maltoheptaose. Hydrogen-bonding interactions (≤ 3.5 Å) are depicted as dashed lines, and Glc residues are labeled from the non-reducing to reducing end.

The crystal structure of the catalytically inactive D498N mutant of SusG with maltoheptaose revealed ligand binding to the CBM58, the active site, and an unexpected surface starch-binding site (SBS) adjacent to the active site [56]. Maltoheptaose binding at CBM58 is 45 Å away and on the opposite side of the protein from the active site, while maltoheptaose bound at the SBS is ~5 Å from a glucose residue of maltoheptaose sitting at the +2 subsite (**Figure 1.3B**). That the SBS is distinct from the active site is demonstrated by the opposite orientation of the maltoheptaose molecules bound at the two sites: the reducing ends of each chain are directed towards each other. This orientation also makes it unlikely that a discrete starch helix can interact with both sites simultaneously.

Both CBMs and SBSs provide a proximity effect by localizing the starch polymer near the catalytic site to enhance catalysis [47, 61]. While both sites display the canonical dual aromatic residue platform that recognizes the shape of the α(1,4)-linked glucan, the manner in which maltooligosaccharide is bound at both sites is different. At CBM58, maltoheptaose is bound with the pitch of the helix parallel to the surface of the protein, whereas maltoheptaose at the surface starch-binding site is directed with the pitch of the helix into the plane of the protein. This difference in binding may differentiate the utility of these sites for starch utilization by *B. thetaiotaomicron*. We initially hypothesized that elimination of the SBS via site-directed mutagenesis would enhance catalysis, allowing a starch polymer better access to the catalytic site. However, the elimination of glycan binding at the SBS did not significantly affect activity on the colorimetric substrate *p* nitrophenol-maltohexaose, or soluble amylopectin and pullulan, but did reduce activity on insoluble corn starch, demonstrating that this site is important

for the recognition of an insoluble substrate. When CBM58 was deleted from the enzyme, the activity of the enzyme against insoluble substrates decreased, but activity towards soluble amylopectin increased by 3-fold. While these data demonstrate that the CBM58 enhances the enzyme's ability to localize to an insoluble starch molecule, *B. thetaiotaomicron* in pure culture cannot grow on insoluble starch such as resistant starch [62]. Therefore in the context of growth, CBM58 may have a different role in starch utilization, perhaps by helping to sequester oligosaccharides released by the active site, or in passing these sugars on to the SusCD complex.

To the best of our knowledge, SusG is the only GH13 with a CBM inserted within the catalytic domain. However, this interrupted domain structure was first noted in rumen *Prevotella ruminicola* GH10 xylanases [63] and more recently in the GH10 xylanases from *B. ovatus* [34] and *B. intestinalis* [64]. While the full-length protein structures of these GH10 enzymes have not been determined we hypothesize that like SusG, the CBMs are simply appended from the catalytic domain with minimal disruption of the GH10 protein fold. In many GH13 enzymes that have one or more CBMs, the CBM is located at the N- or C- terminus and in some cases facilitates dimerization, with the CBM shaping the catalytic cleft of the neighboring polypeptide [57-60]. In SusG, the remote location of CBM58 relative to the active site permits a wider catalytic cleft, a feature that may contribute to the enzyme's broad activity towards amylopectin, pullulan, amylose, maltooligosaccharides, and to a much lesser extent, cyclodextrins [28, 56]. Pullulan degradation by SusG produces panose, and this product reflects the unique ability of SusG's active site to accommodate α(1,6) glycosidic bonds while still solely acting on α(1,4) linkages [56]. This flexible recognition of various α-glucans may

reflect the adaptation of *B. thetaiotaomicron*'s single extracellular GH13 to support growth on a variety of glycan structures that the cell might encounter in a human diet consisting of starch from various sources.

Excluding its CBM58, SusG is most similar in sequence and structure to the *Halothermothrix orenii* AmyA, a member of the GH13_36 subfamily that features enzymes that have amylase activity against α(1,6) containing glucans [65, 66]. The active site of SusG is typical of endo-amylases that hydrolyze endogenous α(1-4) glycosidic bonds using an acid-base double displacement mechanism [67]. In the crystal structure of a catalytically inactive SusG mutant (D498N), maltoheptaose occupies subsites -4 through +3 via an extensive network of direct hydrogen bonding between the O2 and O3 hydroxyls of glucose and polar side chains lining the active site [56] (**Figure 1.3B**). Aromatic stacking between H112 and Glc at the -2 subsite and between Y114 and Glc4 at the -1 subsite likely helps position the chain for efficient hydrolysis. D388 is positioned for nucleophilic attack on the C1 of Glc4; our structure of the active enzyme with acarbose captured this β-glucosyl-D388 covalent intermediate [56]. E431 interacts with the O4 of Glc3, acting as the catalytic acid to protonate the leaving oligosaccharide and activating water to split the β-glucosyl-D388 intermediate. D498 is likely important in stabilizing this intermediate [68].

The products liberated from complete starch degradation by SusG *in vitro* are glucose and maltose [56], yet *B. thetaiotaomicron* does not require SusC or SusD to grow on glucose and maltose. This disparity suggests that SusC and SusD have been maintained throughout this organism's evolution because they are required to import oligosaccharides larger than di- or monosaccharides. It is possible that in the context of

the other Sus proteins at the cell surface, SusG liberates maltooligosaccharides larger than di- or monosaccharides. While the typical size of the glycan that traverses the SusC porin is unknown, growth on maltoheptaose requires SusC and SusD, but not SusG [69], suggesting that maltooligosaccharides at least seven glucose units long can pass through the porin. How SusC works together with SusDEFG to import α-glucans remains a current area of investigation. It has been suggested that the SusC-like proteins from PULs that target chondroitin sulfate [70], xylan [34] and α-mannan [35] also transport larger fragments of their cognate glycan.

The α-glucans that arrive in the periplasm are processed by the neopullulanase SusA and the α-glucosidase SusB to yield glucose, which is imported via an undefined inner membrane transporter [23, 30, 71]. SusA and SusB are essential for starch utilization, and together account for most of the starch-degrading activity from whole-cell lysates compared to SusG alone [28, 31]. The crystal structure of SusB revealed that this GH97 enzyme hydrolyzes maltooligosaccharides via an inverting mechanism, yielding β-D-glucose as a product, which contrasts with the typical retaining mechanism within this glycosidase family of enzymes [72]. SusG has a relatively low affinity for starch (K_m ~ 3.1 mM) compared to SusA (K_m ~ 0.125 mM) [28]. The discrepancy between these two enzyme affinities may reflect SusG's dependence on the starch-binding proteins SusDEF to bring starch within proximity of its active site, or perhaps SusG has evolved to act broadly on a variety of starch substrates at the expense of retaining high specific affinity for one substrate.

SusE and SusF bind starch via multiple carbohydrate binding domains

Initial work by the Salyers lab established that SusCD mediate the majority of starch-binding to the cell surface, and comprise together with SusG the “minimal Sus” required for *B. thetaiotaomicron* growth on starch [26, 27]. Conflicting data suggested that the two remaining lipoproteins encoded within the *sus* operon, SusE and SusF, enhance starch-binding to the cell surface, although their amino acid sequences did not imply a function for these proteins in glycan capture [26]. Genome sequencing later revealed the presence of genes encoding such “putative lipoproteins” within most PULs of *B. thetaiotaomicron* and *B. ovatus* [39], as well as the majority of human gut-adapted Bacteroidetes hinting at a conserved function for these proteins within PUL-encoded Sus-like systems [15]. SusE and SusF belong to one of the most overrepresented protein families within the human gut microbiome, and the enrichment of these types of proteins in this niche underscore their functional importance to these bacteria [73].

The crystal structures of SusE and SusF reveal a multimodular structure comprised of a tandem array of immunoglobulin (Ig)-like folds that bind starch in a manner quite reminiscent of starch-targeting CBMs [74](**Figure 1.4A,B**). SusF contains an N-terminal Ig-like fold proceeded by three β-sandwich domains—Fa, Fb, and Fc—arranged in an S-like configuration (**Figure 1.4B**) [46]. The placement of a proline residue at the midpoint between consecutive domains suggests a lack of conformational flexibility along the length of the protein. SusF is tethered to the membrane via a lipidated cysteine followed by 18 amino acids that create a flexible linker, but the rigidity of the folded protein may help project it off of the membrane to facilitate starch capture. In contrast, SusE has only two starch-binding domains—Eb and Ec, which are similar to the Fb and Fc domains of SusF—and an N-terminal Ig-like domain that was not

observed in the electron density (**Figure 2.4A,C**). A prediction of the SusE N-terminal domain structure suggests it is similar to that of SusF, with a longer, flexible linker between the N-terminal and Eb domains (see model in **Figure 2.8**). The N-terminal domains of SusE and SusF do not contribute to starch-binding [74]. In both proteins, the final two C-terminal domains are packed in a back-to-back arrangement via a hydrophobic interface. The individual starch-binding domains of SusE and SusF share the most structural homology with the X25 domain of *Bacillus acidopullyticus* pullulanase [75] (**Figure 1.4D**). While the crystal structure of this pullulanase with oligosaccharide did not reveal the X25 domain as a CBM, a superposition of the X25 domain with the SusE/F domains reveals conservation of the starch-binding residues, suggesting that this X25 may bind glycan (**Figure 1.4E**).

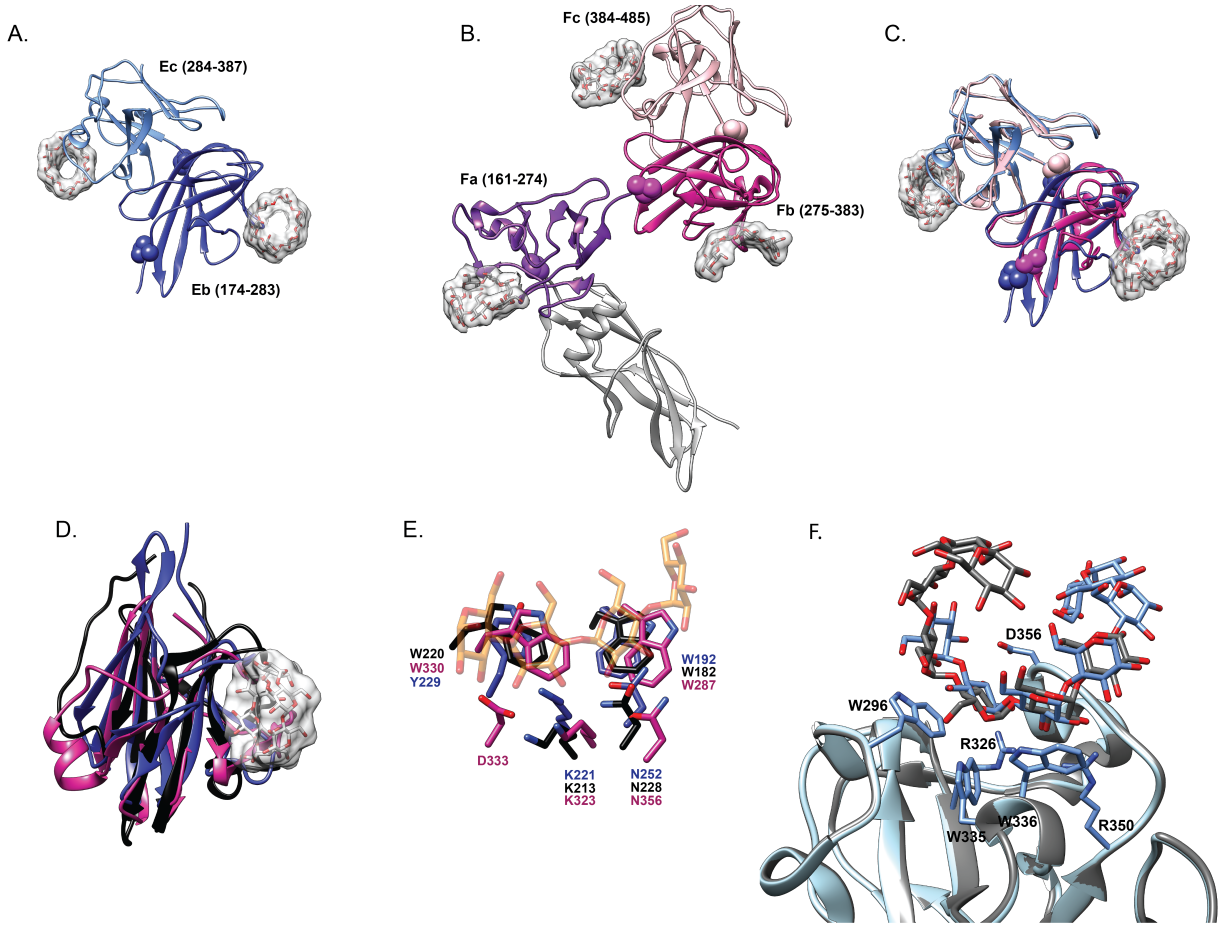

Figure 1.4. Structures of the SusE and SusF proteins

A. Structure of SusE with bound α-cyclodextrin (PDB 4FEM), with the starch-binding domains Eb and Ec in different colors. Proline residues between the domains are highlighted as spheres. B. Structure of SusF with bound maltoheptaose (PDB 4FE9), with the starch-binding domains Fa, Fb and Fc in different colors. Proline residues between domains are highlighted as spheres. C. Overlay of the Eb/Ec and Fb/Fc domains of SusE and SusF, colored as in panels A & B. D. Superposition of the Eb domain (blue), Fb domain (pink) and the X25 domain (black, residues 161-270 of PDB 2WAN) from the *Bacillus acidopullulyticus* pullulanase [75]. E. Close-up of the starch-binding sites in Eb and Fb from the overlay in panel D, demonstrating that these residues are conserved within the X25 module of the pullulanase (PDB 2WAN). Residues and labels are colored as in Panel D, and the portion of the α-cyclodextrin bound to Eb is displayed as transparent orange and red sticks. F. Overlay of the two positions that maltoheptaose occupied at the Ec binding site of SusE (PDB 4FCH), demonstrating how a longer single-helical stretch of amylose could be accommodated.

SusE and SusF are not described as CBMs as this would conflict with the definition of a CBM as a domain appended to a carbohydrate-active enzyme. However, although SusE and SusF are independent proteins physically separate from the α-amylase SusG, they may provide a similar functionality to a traditional CBM in the context of the Sus. The design of the Sus outer membrane protein complex, whereby glycan capture and carbohydrate degradation is spread over multiple polypeptides, is vaguely reminiscent of the cellulosomal architecture [27]. Cellulosomes are multiprotein structures comprised of enzymes and some distinct CBMs that assemble into a complex for efficient cellulose deconstruction [76, 77] and a similar system for starch hydrolysis has recently been identified in the Firmicute *Ruminococcus bromii* [16]. The cellulosome is held together by a system of complementary protein-protein interaction domains called cohesins and dockerins, motifs that are not found in the Sus. However, Salyers and others have demonstrated that proteins within Sus and Sus-like systems also interact [27, 78, 79].

The five starch-binding domains shared between SusE and SusF display a similar starch-binding architecture featuring two aromatic amino acids that provide a hydrophobic interface for α-glucan binding, plus additional residues that hydrogen-bond with the hydroxyl groups of the glucose residues to stabilize the interaction (Table 2). Subtle differences in the arrangement of glycan-binding residues likely explains the somewhat different affinities of each domain for maltoheptaose versus α-cyclodextrin (Table 2) [74]. All of the SusEF domains show weaker binding for glucosyl-α(1,6)-maltotriosyl-α(1,6)-maltotriose (GMM), a linear oligosaccharide of pullulan, compared to maltoheptaose suggesting that SusE and SusF have not been adapted for α(1,6)

recognition. The most divergent domain between SusE and SusF is Ec, which displays the highest affinity for maltoligosaccharides. In this domain a loop within this binding site defined by residues 353-357 caps one end of the α-glucan binding site. In this crystal structure, maltoheptaose was shared across a crystallographic symmetry axis between chain A from one asymmetric unit and chain B from another. Superposition of these chains simulated a 10-glucose long maltooligosaccharide that is wound up and over this capping loop (**Figure 1.4F**). We suggest that this binding site has been adapted to preferentially recognize single-helical regions of starch, a feature that may help the Sus complex to target partially denatured regions of starch that could be more easily hydrolyzed by SusG.

Interestingly, despite the tandem arrangement of starch-binding domains in both SusE and SusF, only the domains of SusE synergize and display enhanced binding to insoluble corn starch via an avidity effect; SusE mutants with a single functional domain display greatly decreased binding [74]. In comparison, the native full length SusF protein binds insoluble cornstarch nearly as well as a site-directed mutant possessing only a functional Fc domain (i.e., both the Fa and Fb sites were ablated) [74]. Thus the individual domains of SusF do not appear to enhance overall protein binding to starch, but rather the Fc domain is largely responsible for the SusF starch-binding affinity. Although there are distinct structural and biochemical differences between SusE and SusF, it is not yet clear what functional differences these proteins have in the context of the Sus or *B. thetaiotaomicron*'s ability to utilize starch in the gut.

Differentiating the roles of the SusDEFG starch-binding sites in starch utilization

Among SusDEFG there are eight starch-binding sites present on the surface of *B. thetaiotaomicron*, yet the roles of these sites within the Sus are distinct. The most critical starch-binding protein is SusD, as an in-frame deletion of *susD* (Δ*susD*) results in the loss of growth on starch [42]. In this mutant, transcription of *susEFG* occurs at wild type levels, supporting the hypothesis that the growth defect is due to the loss of SusD, and presumably, the loss of starch-binding to the cell surface conferred by SusD. At the time that we created the Δ*susD* strain, we did not know that SusEFG also possessed starch-binding domains. To determine if the loss of growth on starch in the Δ*susD* strain was due to the loss of starch-binding by SusD, we compared the growth of the Δ*susD* to a Δ*susD* strain complemented with the allele for SusD* (Δ*susD::susD**),

a site-directed mutant of the SusD binding site that eliminates starch-binding *in vitro* [69]. The Δ*susD::susD** cells grow on starch (5 mg/ml) when *sus* transcription is activated by the addition of a small amount of maltose (0.5 mg/ml) to the media. Furthermore, *ΔsusD::susD** cells can grow on maltoheptaose with wild-type growth kinetics and without the need for maltose induction in contrast to Δ*susD* cells. We concluded from these experiments that the presence of SusD was essential for growth on starch, although a SusD that also binds starch facilitates more efficient growth without the need for transcriptional activation by maltose. Indeed, quantification of *sus* transcript from both wild-type and SusD* expressing cells exposed to various concentrations of maltooligosaccharides revealed that the SusD* cells required 100-1000 fold higher concentrations of glycan than wild-type cells in order to achieve wild-type transcriptional activation. The role of SusD in sugar sensing is likely indirect; we speculate that starch/maltooligosaccharide binding by SusD enhances the rate of import through SusC, leading to an accumulation of sugar in the periplasm that is sensed by SusR resulting in *sus* transcriptional activation. These data from the SusD* growth experiments support a role for SusD in starch utilization that hinges upon its interaction with SusC. The physical presence of SusD may help stabilize SusC, or otherwise permit the assembly of a larger complex containing the rest of the Sus proteins.

Unlike SusD, starch-binding by SusEF does not contribute to α-glucan sensing by inducing *sus* expression, although the *sus* transcriptional response is somewhat diminished when a deletion of both SusEF from the cell surface is combined with mutations in either the SusG SBS *(*Δ*susG*^{SURF}) or CBM58 *(*Δ*susG-CBM58*) [69]. Rather, SusEF influence the growth of *B. thetaiotaomicron* in a substrate-dependent manner.

Growth on high molecular weight, highly branched maize amylopectin is impeded in cells with a combined deletion in SusEF and one of the SusG starch-binding sites [69]; this growth defect was not observed on potato starch that has a lower molecular weight. This observation lead to the hypothesis that the multiple starch-binding sites of SusEFG aid in acquiring high molecular weight starch species through the thick capsule layer of *B. thetaiotaomicron*. Like most human gut-adapted *Bacteroides* species, *B. thetaiotaomicron* produces a polysaccharide capsule several hundred nanometers thick [80], which likely protects the cell from the host immune response, but could impose a diffusion barrier to nutrients that must reach the cell surface. While the Sus proteins do not protrude above the capsule layer, they seem to aid in the capture of starch through this extracellular matrix, as a Δ*susEFG*^{SURF} or Δ*susEFG-CBM58* mutant in an acapsular strain of *B. thetaiotaomicron* does not display a growth defect on maize amylopectin [69].

The specialized roles of SusDEFG in starch utilization are apparent *in vivo* as well. To test how the individual Sus proteins adapt the cell to scavenge starch in the host intestinal tract, germ-free mice were co-colonized with wild-type, Δ*susC,* and either Δ*susD::susD*,* or Δ*susEFG*^{SURF} *B. thetaiotaomicron [69]*. Mice were fed a diet high in resistant starch and additionally colonized with or without *Ruminococcus bromii*, a resistant starch degrading species that may cross-feed maltooligosaccharides to *B. thetaiotaomicron* [7, 62]. The Δ*susC* and Δ*susD::susD** mutants were outcompeted by the wild-type strain in the presence or absence of *R. bromii*. Interestingly, the Δ*susEFG*^{SURF} mutant fared as well as wild-type *B. thetaiotaomicron* when *R. bromii* was absent, but was quickly outcompeted when *R. bromii* was also present. Here, *R. bromii*

may increase the abundance of larger maltooligosaccharides that require further processing prior to transport. These data suggest that *B. thetaiotaomicron*'s multiple starch-binding sites have evolved to optimize nutrient acquisition within the competitive polymicrobial environment of the colon.

The Sus complex dynamically assembles in the presence of starch

The observed cooperation between starch-binding sites during starch degradation and import implies that the Sus proteins are working closely together to optimize starch utilization. Salyers and colleagues demonstrated that SusCD interact, and that SusE may also interact with this complex [27]. Additionally, both SusEF are less sensitive to proteolytic degradation when expressed together on the cell surface, suggestive of complex formation [27]. In the Bacteroidetes *Capnocytophaga canimorsus*, affinity purification of the SusC-like transporter GpdC, required for host N-glycan utilization, resulted in the co-purification of the SusD-like protein GpdD, the surface glycosylase GpdG, and a periplasmic sialidase [78]. This suggests that interactions among Sus-like proteins may be conserved across different glycan utilization systems within the Bacteroidetes. However, the nature of these protein-protein interactions has not been defined.

The dynamic movement of SusG on the cell surface has been captured by single-molecule fluorescence imaging in live *B. thetaiotaomicron* [79]. In these experiments, a mutant of SusG was created by replacing the CBM58 with a HaloTag (HT) protein, which was fluorescently labeled by the dye tetramethyl rhodamine [81, 82]. We tracked the diffusion of this tagged SusG (SusG-HT) in live *B. thetaiotaomicron* cells

in the presence of glucose and starch. Under all conditions, we observed both freely diffusing and slow-moving SusG-HT. We hypothesize that these slow-moving species occur due to the interaction of SusG with other Sus proteins. In addition, the net movement of SusG-HT decreased in the presence of amylopectin compared to glucose, likely due to the interaction of SusG-HT with the polymer. However, SusG-HT mobility in the presence of starch increased in both Δ*susD* and Δ*susEF* cells, presumably because SusG was not able to associate with these proteins [79]. We believe these data suggest that SusG dynamically associates with other Sus proteins, resulting in a slow-moving population of molecules, and that during growth on starch the polysaccharide may effectively "crosslink" the Sus proteins, promoting or stabilizing their interactions.

Variation in the Sus across the Bacteroides

The structure of the *sus* operon of *B. thetaiotaomicron* is not completely conserved as there are several variations of predicted starch-targeting PULs among other well-studied *Bacteroides* species (**Figure 1.5**). In particular, the number of SusE/F homologs, and conservation of the SusG protein varies extensively. For example, *Bacteroides fragilis* encodes one SusE, and a SusG homolog, both of which are longer than their homologs in *B. thetaiotaomicron*, and have limited identity over the length of the polypeptide. In addition, many predicted Sus PULs do not include obvious *susA* or *susB* genes within the same operon. How these variations in operon structure, protein sequence (and hence structure) affect starch utilization in these organisms is unknown. However, this comparison highlights that the proteins encoded by the *susC*

and *susD* genes are the most well-conserved, underscoring their central function in glycan uptake.

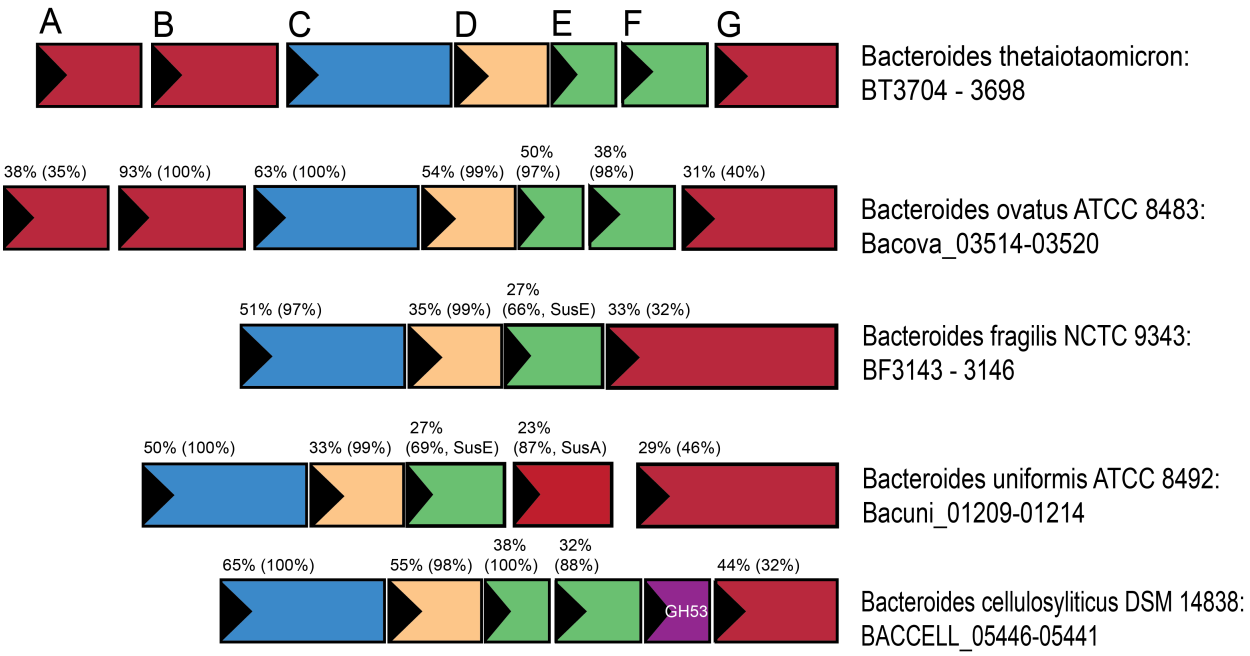

Figure 1.5. Sus operon structure across *Bacteroides* species

Selected *sus* operons from the type strains of several human gut *Bacteroides* species are displayed, identified via conservation of the operon structure surrounding a *susG* homolog. Numbers displayed above each gene indicate the percent identity of the encoded protein and in parentheses the percent coverage of the match to the homologous protein from *B. thetaiotaomicron*. For example, 63% (100%) above the *susC* homolog from *B. ovatus* indicates that the SusC homolog from B. ovatus is 63% identical to the *B. thetaiotaomicron* SusC and that this match covers 100% of the protein sequence.

Structural basis for TonB-dependent transport in Sus-like systems

At the core of outermembrane glycan capture is the SusCD complex, but the molecular details of the relationship between starch binding by SusD and starch import through SusC are unclear. Recently, the crystal structures of two distinct SusCD-like complexes were solved and the details of their interaction have made major contributions to a general model of nutrient uptake by homologous SusCD-like complexes [83]. The putative peptide-targeting PUL BT2261 – 2264 encodes a SusC-like TonB-dependent transporter (BT2264), SusD-like substrate-binding protein (BT2263) and two additional lipoproteins composed of Ig-like folds akin to SusE and SusF (BT2261-2) (**Figure 1.6A**). All four of these proteins were found to interact tightly within a octomeric protein complex that was purified in a single peak from anion exchange. The levan-targeting PUL encodes a SusC homolog (BT1763) and SusD homolog (BT1762) that form a dimeric protein complex that was was purified over a nickel column using a C terminal his-tag on BT1762 (**Figure 1.6B**). In contrast to the tetrameric BT2261 – 2264 complex, the BT1762 – 2263 dimer does not include the levan-binding SusE-like homolog (BT1761) or the cell surface-exposed GH32 enzyme (BT1760), despite their co-regulated expression and membrane colocalization.

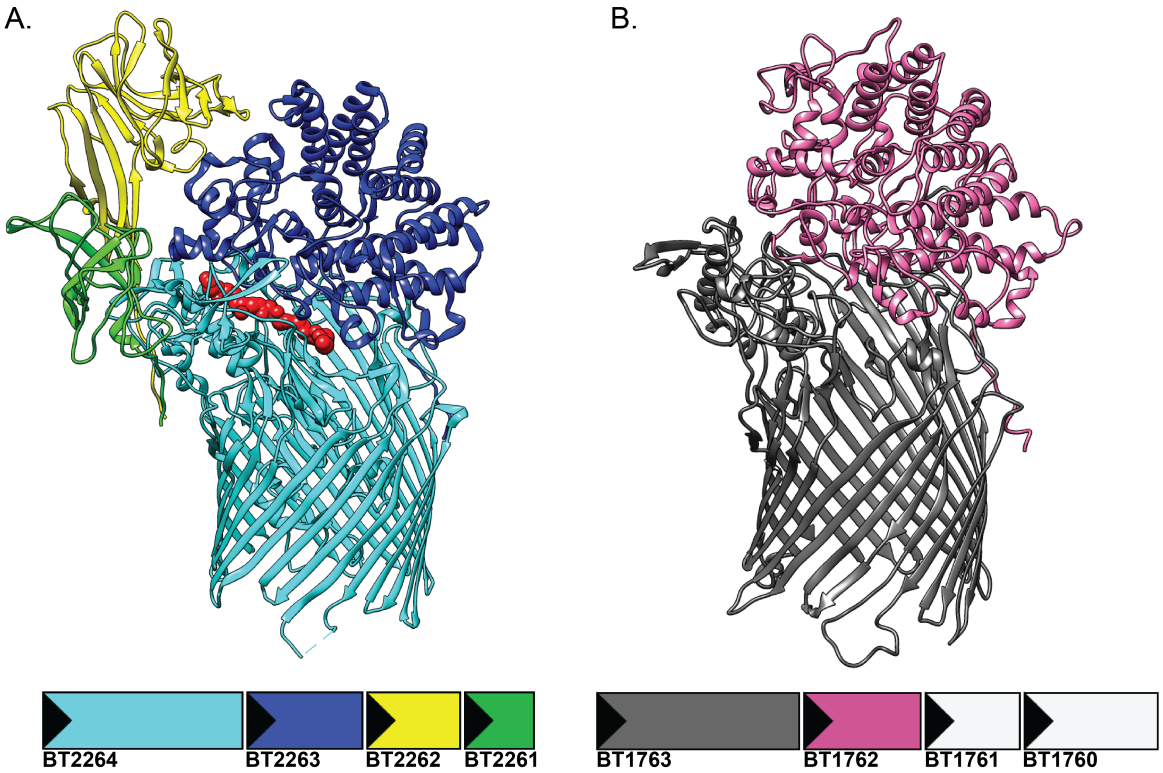

Figure 1.6. Architecture of SusCD-like complex assembly

A. Crystal structures and locus organization of the putative peptide-targeting complex BT2261 – 2264 (PDB 5FQ6) and B. the levan-targeting complex BT1762 – 1763 (PDB ID 5T3R) [83]. Each protein is colored to match the corresponding gene. The peptide within BT2261 – 2264 is colored in red and is seen sandwiched between the substrate binding site of SusD-like BT2263 and the plug domain of SusC-like BT2264. The plug domain of BT1763 is missing from the structure because it was sensitive to proteolytic cleavage following purification.

The SusC-like homologs BT2264 and BT1763 form a β-barrel composed of 22 β-sheets and 11 extracellular loops, similar to previously studied TonB-dependent transporters (**Figure 1.6A,B**) [83, 84]. A plug domain embeds itself inside the periplasmic side of BT2264 and occludes the movement of substrates through the lumen of the barrel. This domain is thought to undergo the largest TonB-mediated conformational changes that lead to substrate translocation through the transporter. The plug domain of BT1763 was sensitive to proteolytic cleavage during purification and was not present in the structure, supporting the idea that the plug domain is completely removed upon substrate transport, in contrast to what is seen during iron transport in TBDTs. In both cases, the SusD-like homologs interact in the same configuration with their cognate SusC-like transporters. BT2263 and BT1762 are thought to behave like sealed lids by directing their substrate binding sites to a solvent inaccessible compartment inside BT2264 and BT1763, respectively.

Molecular dynamic simulations suggest that despite the large interaction surface between SusD-like BT2263 and the TBDT BT2264, this peptide-binding protein may open over the pore in a hinge-like motion that allows its binding site to sample the extracellular space above BT2264 and configure a peptide into the transporter [83]. These simulations demonstrate that the iterative dynamics in the BT2263 – 2264 complex may support a processive mechanism for substrate binding and transport. Furthermore, this strategy may reflect how Sus-like systems have adapted to efficiently transport repetitive fragments of larger polysaccharides that seem to be the major carbon source targeted by the Bacteroidetes. Although these data have provided critical mechanistic insights into the initial steps of substrate translocation across the outer

membrane, to what extent the plug domain changes conformation to allow for transport is still unknown and may represent a necessary adaptation for glycan uptake by these systems.

Summary and Chapter Outline

The starch utilization system of *B. thetaiotaomicron* is composed of eight genes, five of which encode proteins that localize to the outer membrane of the cell where starch is first encountered. These proteins collectively bind and degrade large starch polysaccharides so that smaller oligosaccharides can be shuttled into the cell for further hydrolysis and energy harvest. Delineation of the biochemical and structural features of the individual Sus proteins has facilitated the development of a working model for how the Sus proteins, and likely homologous proteins from other Sus-like systems within the Bacteroidetes, interact to metabolize carbohydrate nutrition (**Figure 1.7**). In this model, the SusCD proteins are key for initial starch sensing, and together promote the efficient uptake of maltooligosaccharides. SusC and SusD likely interact frequently as the essential unit for glycan uptake, while the interactions of these proteins with SusEFG may be more dynamic. SusEF as well as the starch-binding sites within SusG support starch binding at the cell surface through the polysaccharide capsule. The redundancy of cell surface starch-binding sites likely enhances the capture of dietary starch, and maltooligosaccharides generated by other species in the gut. Notably, the dynamic assembly of the Sus proteins may enhance starch capture by allowing each protein additional degrees of freedom for optimal starch-binding. Lastly, seminal structural studies of the SusCD-like homologs have provided a molecular model for the

mechanics of these two central PUL-encoded proteins, accelerating our knowledge of the least-well understood aspect of polysaccharide translocation across the outer membrane.

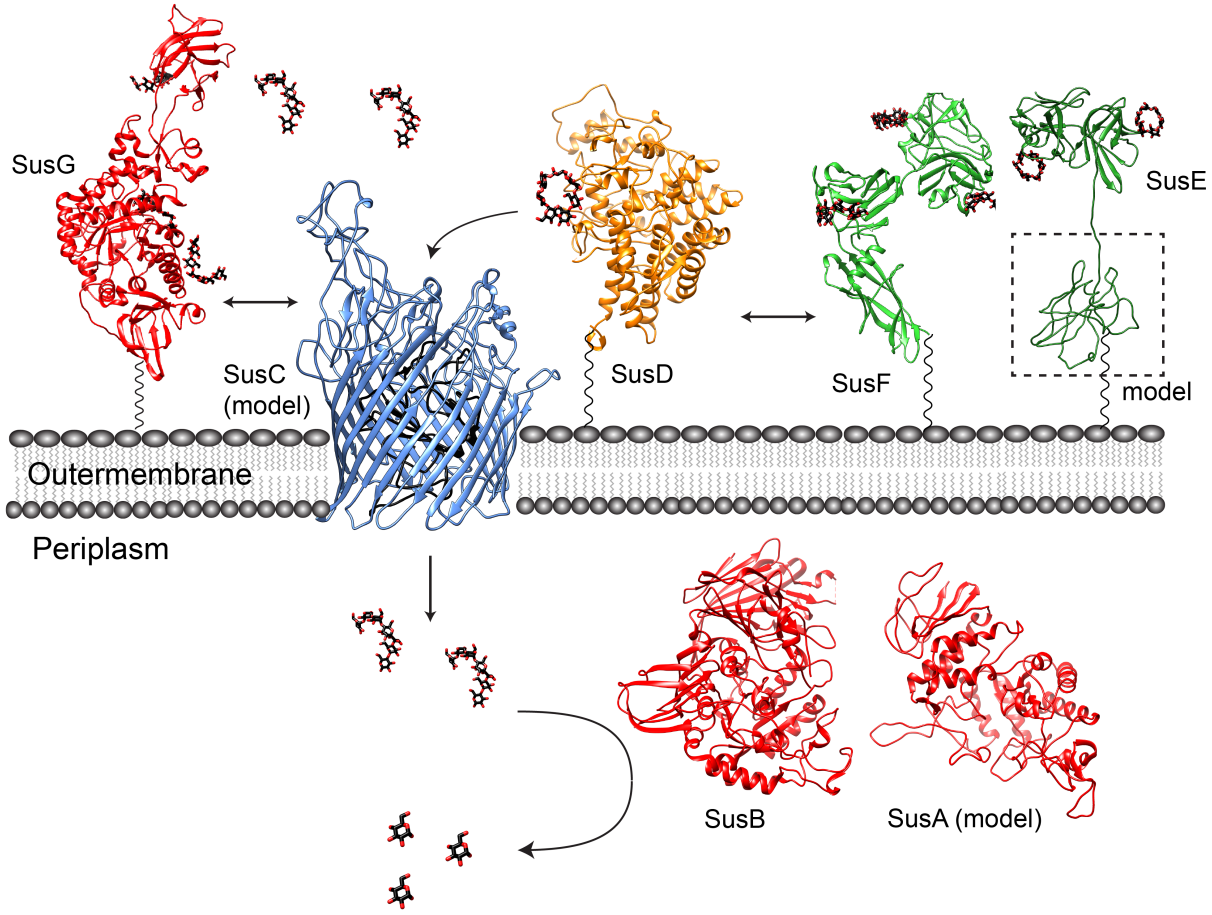

Figure 1.7. Sus protein structures and model of dynamic assembly

Ribbon diagrams of the crystal structures and homology models for the seven Sus proteins involved in starch utilization in *B. thetaiotaomicron*, colored as in **Figure 1.1**. The flexible amino acid sequences that link SusDEFG to the membrane are depicted as a black wavy lines, as these residues were not resolved in the crystal structures. SusG (PDB 3K8L) dynamically interacts with SusD (PDB 3CK9) and SusC (ITASSER structure prediction [85, 86]). SusE (PDB 4FEM; ITASSER prediction of the N-terminal domain, with modeling of the linker sequence) and SusF (PDB 4FE9) may directly interact with each other and with the SusCD complex, as suggested by Salyers [27]. Maltooligosaccharides are transported through the SusC TonB-dependent transporter where they are further hydrolyzed to glucose by the action of SusB (PDB 2JKA) and SusA (model from Modpipe [87] using template PDB 3DHU). Maltooligosaccharides and glucose are depicted as black and red sticks.

The Sus is a model system for glycan uptake by mammalian gut Bacteroidetes, and the repertoire of Sus-like systems encoded within the genomes of these organisms dictates their glycan utilization profile [40, 88]. As the field moves toward a molecular-level understanding of the organization and function of other Sus-like systems, we will see how this basic paradigm as outlined for the Sus of *B. thetaiotaomicron* has been adapted for the capture of diverse glycans from the gut environment.

In Chapter II we describe new functional roles for the outer membrane lipoprotein SusE, demonstrating that SusE but not SusF can compensate for the loss of starch-binding by SusD during growth on maltooligosaccharides. This compensatory effect is not dependent on the starch-binding sites of SusE, suggesting that the assembly of SusCDE is important for starch uptake. Additionally, the starch-binding sites in SusDEF are shown to facilitate the uptake of longer maltooligosaccharides through the transporter SusC.

In Chapter III we employ super resolution imaging and single molecule tracking to determine that unlike SusG that diffuses through the outermembrane, SusE and SusF remain stationary. SusC, SusD, and SusE are shown to interact by co-immunoprecpitation, however SusE and SusF are shown to remain immobile even when the outer membrane members of Sus are genetically deleted. These observations provide evidence for protein complex formation in which mobile enzymes assemble around immobile glycan-binding proteins.

In Chapter IV, we assess the roles of the outermembrane proteins in the Xyloglucan Utilization Locus (XyGUL) of *B. ovatus*. XyGUL outermembrane glycan-binding proteins display a hierarchy for efficient xyloglucan uptake. The structural and

phenotypic characterization of a cell surface GH9 enzyme provides evidence for a unique mechanism of XyG uptake in the *B. ovatus* and may hint at unappreciated functions for PUL-encoded cell surface enzymes.

Finally, in Chapter V we characterize the role of the glycan-binding proteins and sites within the Mixed-Linkage β-glucan Utilization Locus (MLGUL) of *B. ovatus*. Unlike in other PUL, the SusD-like homolog of MLGUL, $SGBP_{MLG}$-A, requires a functional glycan-binding site to support growth on the polysaccharide. Additionally, we analyze the preferences for oligosaccharides by the MLGUL SusE-like homolog, $SGBP_{MLG}$-B.

Chapter II: SusE facilitates starch uptake independent of starch binding in *B. thetaiotaomicron*

Notes

This chapter was reprinted and modified with permission from Foley, M.F., Martens, E.C., Koropatkin, N.M. SusE Facilitates Starch Uptake Independent of Starch Binding in *B. thetaiotaomicron*. *Mol Microbiol.* 2018 March 12.

Abstract

The *Bacteroides thetaiotaomicron* starch utilization system (Sus) is a model system for nutrient acquisition by gut Bacteroidetes, a dominant phylum of gut bacteria. The Sus includes SusCDEFG, which assemble on the cell surface to capture, degrade and import starch. While SusD is an essential starch-binding protein, the precise role(s) of the partially homologous starch-binding proteins SusE and SusF has remained elusive. We previously reported that a non-binding version of SusD (SusD*) supports growth on starch when other members of the multi-protein complex are present. Here we demonstrate that SusE supports SusD* growth on maltooligosaccharides, and determine the domains of SusE essential for this function. Furthermore, we demonstrate that SusE does not need to bind starch to support growth in the presence of SusD*, suggesting that the assembly of SusCDE is most important for maltooligosaccharide

uptake in this context. However, starch binding by proteins SusDEF directs the uptake of maltooligosaccharides of specific lengths, suggesting that these proteins equip the cell to scavenge a range of starch fragments. These data demonstrate that the assembly of core Sus proteins SusCDE is secondary to their glycan binding roles, but glycan binding by Sus proteins may fine tune the selection of glycans in the environment.

Introduction

The human gut microbiota performs critical tasks that promote host health and development [1-4] including the mutualistic break down of complex carbohydrates (fiber) from our diet, [2, 89]. Dietary fiber processing begins at the bacterial cell surface via the action of one or more glycoside hydrolases or polysaccharide lyases that liberate smaller oligosaccharides that are harvested by the same or neighboring bacterial species [90]. For example, organisms in the gut such as *Ruminococcus bromii* process resistant starch into small fragments that can then cross-feed other species [62]. However, other bacteria such as *Bacteroides thetaiotaomicron* may employ a more "selfish" mechanism for the processing of specific glycans such as yeast cell wall a-mannan in which the bacterium breaks α-mannan into fragments that it can select for uptake, limiting cross-feeding [35]. Understanding the molecular mechanisms employed by gut microbes to utilize dietary carbohydrates can foster the development of dietary strategies to manipulate the microbiota and improve health.

The Gram-negative Bacteroidetes are abundant members of the intestinal ecosystem in part due to their ability to efficiently acquire carbohydrates [15, 91]. The

Bacteroidetes employ complexes of proteins, termed Sus-like systems, that localize to the cell surface and act in concert to bind, degrade, and import glycans. Sus-like systems are encoded in polysaccharide utilization loci (PUL), gene clusters that are transcriptionally activated in response to a distinct carbohydrate substrate [15]. Some Bacteroidetes dedicate ~ 20% of their genomes to encoding PUL [39]. All PUL have at least one pair of proteins that share homology to SusC, a putative TonB-dependent transporter, and SusD, a starch-binding protein, from the Starch utilization system (Sus) of *Bacteroides thetaiotaomicron [15, 92]*. Detailed biochemical studies of the glycolytic machinery and glycan-binding proteins from PUL that target pectin, β-glucan, xyloglucan, chitin, and xylan, among others, have helped elucidate a general model of these systems by which carbohydrates are initially degraded at the cell surface and oligosaccharides are selected and imported into the cell via a SusCD-like complex [33, 34, 93-95].

The *B. thetaiotaomicron* Sus is an eight-gene locus that encodes five outermembrane proteins, SusCDEFG, involved in starch uptake [24] (**Figure 2.1**). X-ray crystallographic structures of SusDEFG revealed eight distinct starch-binding sites among the four proteins [42, 56, 74]. SusG is the sole extracellular Sus enzyme and is required to process starch into maltooligosaccharides that are imported via the putative TonB-dependent transporter (TBDT) SusC [28, 29]. In the periplasm, maltooligosaccharides are depolymerized by SusA, a neopullulanase, and SusB, an α-glucosidase [30, 31, 72]. The disaccharide maltose is recognized by the regulatory protein SusR, and activates *sus* transcription [25].

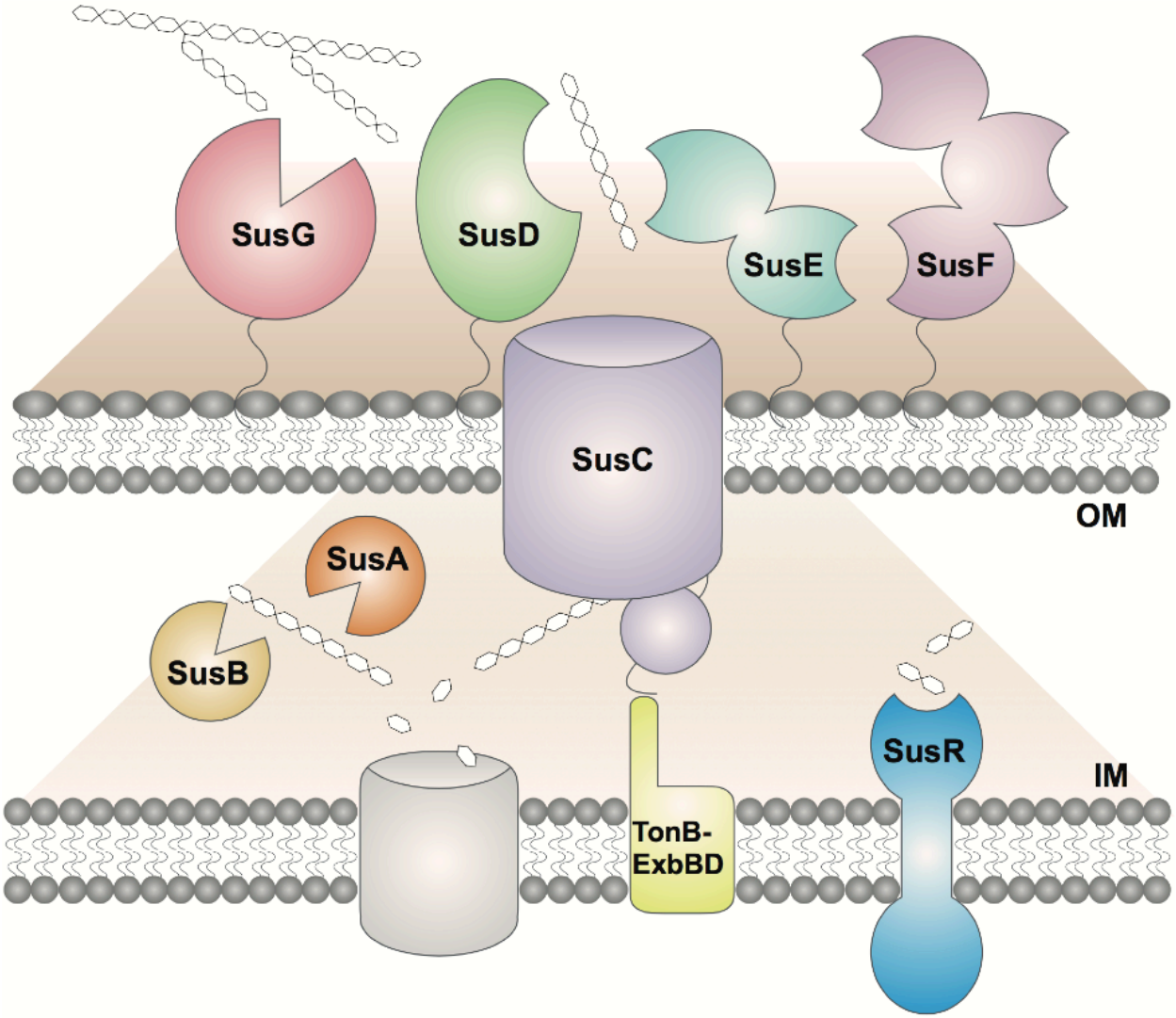

Figure 2.1. Overview of the Starch Utilization System (Sus) in *B. thetaiotaomicron*

Starch is bound by the starch-binding lipoproteins SusDEF and the α-amylase SusG cleaves the polysaccharide to generate maltooligosaccharides that can be internalized by the TonB-dependent transporter SusC. The neopullulanase and α-glucosidase SusA and SusB, respectively, process the maltooligosaccharides to glucose that is introduced to the cytoplasm through an unknown transporter. The inner membrane-spanning protein SusR senses the disaccharide maltose and subsequently drives transcription of the *sus* locus.

One important role of SusD during glycan transport is the capture of maltooligosaccharides for efficient import via SusC, which leads to increased levels of these sugars in the periplasm and triggers transcriptional activation of *sus* via SusR [69]. We have demonstrated that cells with a starch- binding-deficient allele of SusD, termed SusD*, are much less sensitive to the presence of maltooligosaccharides in the environment, requiring 100-1000x more sugar to achieve maximum transcription of the *sus* operon. SusD* expressing cells also lag longer than SusD-expressing cells when cultured on starch or maltooligosaccharides but not on glucose. However, the addition of small amounts of maltose, which upregulates *sus* and but does not require Sus for uptake, partially relieves the lag seen with SusD* cells. These data suggested to us that the single starch-binding site of SusD is not required during steady-state growth of *B. thetaiotaomicron* on starch and maltooligosaccharides (i.e. when the *sus* operon is already upregulated) implying that SusE and SusF may provide substrate binding during transport [69]. Seminal work performed by Salyers and colleagues demonstrated that SusC and SusD interact and that SusEF may modulate the SusCD interaction [27]. The recent structures of two SusCD-like complexes from *B. thetaiotaomicron* revealed that the SusD-like protein sits on top of the SusC-like transporter, with the putative substrate-binding site facing the channel [83]. In one of these complexes, multidomain PUL-encoded proteins resembling SusE and SusF co-purified with the transporter, suggesting that other PUL-encoded proteins may interact with the SusCD-like complex. Co-immunoprecipitation of Sus-like proteins that target host sugars such as heparin, heparin sulfate, and other *N*-linked glycoproteins also demonstrate that the cell-surface glycan-binding proteins physically interact [78, 96].

In this study, we investigated how the SusDEF lipoproteins and their starch-binding sites contribute to starch and maltooligosaccharide utilization by *B. thetaiotaomicron*. Using targeted mutations in SusDEF, we demonstrate that only SusE can compensate for the loss of starch binding by SusD (SusD*) during growth on maltoheptaose. Most strikingly, we observed that the starch-binding function of SusE is not required to support growth of the SusD* mutant on maltoheptaose. Finally we reveal how the SusDEF starch-binding lipoproteins drive the preference of the Sus complex for maltooligosaccharides of different size ranges. Together, these results show that SusD and SusE do not require starch-binding sites to direct maltooligosaccharides through SusC, yet SusDEF have a profound impact on the selection of maltooligosaccharides. These insights enhance our understanding of polysaccharide substrate capture by *B. thetaiotaomicron* and, more broadly, carbohydrate utilization in the Bacteroidetes.

Results

SusE compensates for the loss of the SusD starch-binding site during growth on maltoheptaose

Enzymatic cleavage by the SusG enzyme liberates starch-derived maltooligosaccharides that are the primary substrates of the remaining four Sus outermembrane proteins. To address the roles of SusE and SusF during growth on maltoheptaose, a model maltooligosaccharide, we created a series of genetic deletions and point mutants in the genes encoding these proteins and recombined these back into the *sus* operon via allelic exchange in a Δtdk strain of B. thetaiotaomicron, referred

to throughout as the wild-type strain. Alleles of *susDEFG* that have been mutated to abolish starch binding are labeled as Sus* mutants, and all have been previously characterized (**Table 2.1**). All strains used in this study or created within our previous studies are listed in Table S1. To determine how the SusEF proteins in our strains affect growth on starch or maltooligosaccharides, cell growth was monitored anaerobically in a 96-well plate reader. For each growth experiment, all strains were back-diluted 1:200 from minimal media with glucose or maltose, as noted, into both the test substrate (starch or maltooligosaccharide) as well as a glucose control. All of the 24 mutants reported in this study grew as the parent strain on glucose (**Figure 2.2**), and achieve exponential phase at nearly the same time, verifying that all strains in each set of growth experiments were started at the same O.D. Therefore despite the limitations of the plate reader in resolving growth at low O.D., we feel confident that the dramatic differences in lag that we see with some of our phenotypes reflects biological differences among the strains and is not an artifact of our experimental set-up.

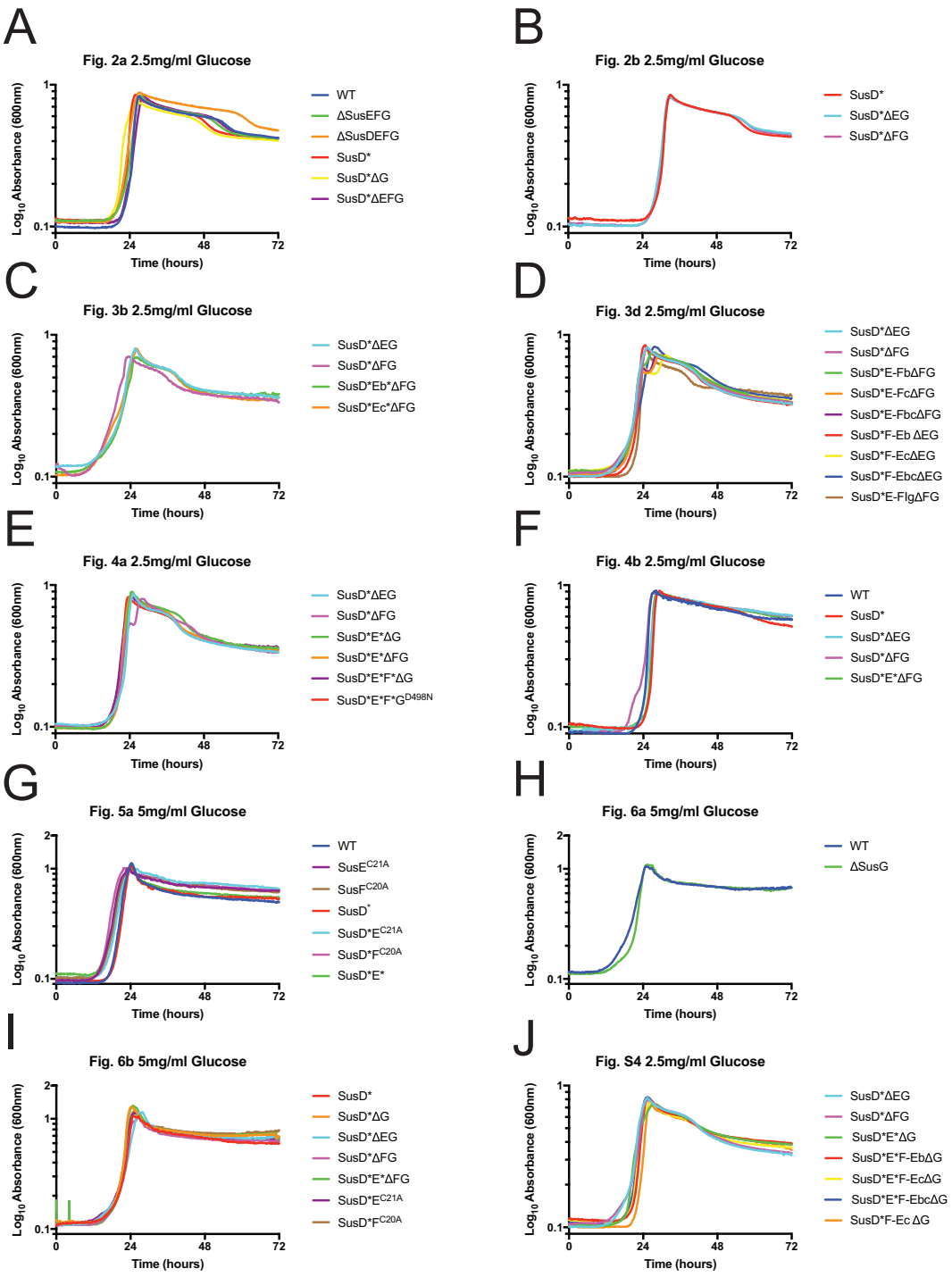

Figure 2.2. Compiled growths on glucose

(A-J) Average growth curves of *B. thetaiotaomicron* strains on glucose. Panels are organized chronologically through the main text, and labeled according to the main figure that they correspond with. Note that the concentration of glucose used matches the concentration of the substrate used in the corresponding main text figure.

Table 2.1. Mutant Sus alleles used in this study

Mutant allele	Mutations (by protein residue number)	Effect	Reference
SusD*	W98A/N101A/W320A	No starch binding	[69]
SusEb*	W192/K221A/Y229A/N252A	No starch binding at the Eb domain	[74]
SusEc*	R326A/W336A/R350A	No starch binding at the Ec domain	[74]
SusE*	W192/K221A/Y229A/N252A/ R326A/W336A/R250A	No starch binding at the Eb or Ec domains	[74]
SusF*	W177A/K208A/W222A/D231A/W287A/K323A/N356A/W396A/W442A/R456A	No starch binding at the Fa, Fb or Fc domains	[74]
$SusG^{D49N}$	D498N	Catalytically inactive (nucleophile mutant) SusG	[56]
$SusE^{C21A}$	C21A	Periplasmically localized SusE	[74]
$SusF^{C20A}$	C20A	Periplasmically localized SusF	[74]

As previously observed, *B. thetaiotaomicron* cells expressing SusD* grow on maltoheptaose despite an extended lag compared to the wild-type parent strain (**Figure 2.3A**). This lag is associated with reduced transcriptional activation of the *sus* operon, presumably due to inefficient maltooligosaccharide uptake and reduced glycan levels in the periplasm [69]. Growth on maltoheptaose is not dependent on extracellular processing by the surface enzyme SusG as SusD*DG cells grow the same as the SusD* strain. Additionally, cells lacking SusDEFG cannot grow on maltoheptaose, demonstrating that SusC alone is not sufficient to support growth on maltoheptaose. In order to better resolve the apparent lag between the wild type and SusD* strain, a second growth experiment with biological triplicate cultures on 2.5mg/ml maltoheptaose (2.17 mM) was performed in culture tubes and the $O.D._{600}$ was assessed in a spectrophotometer to resolve lower $O.D._{600}$ readings (**Figure 2.3B**). The growth of these strains revealed a biphasic curve with an early exponential expansion containing similar specific growth rates at $O.D._{600}$ = 0.01 between wild-type (0.262 ± 0.13) and SusD* (0.271 ± 0.05). A larger, second exponential phase revealed an increased growth rate at $O.D._{600}$ = 0.2 for wild-type (0.678 ± 0.07) compared to SusD* (0.399 ± 0.09). SusD* displays an initial lag before growth as well as a slower second exponential growth rate relative to wild-type, and both contribute to the apparent lag observed in the plate reader. The precise difference between growth rates in the culture tubes vs the plate reader is unclear, but not entirely unexpected when growth conditions change. We suspect that evaporation in the plate reader set-up may contribute to these effects. The growth defects seen in SusD* are likely representative of the apparent lag phenotypes observed in other strains in this study.

The ability to grow on maltoheptaose is abolished when the SusD* allele is combined with loss of SusEFG, hinting at a compensatory role for SusE and/or SusF. Upon further analysis, individual deletions of *susEG* and *susFG* revealed that the loss of SusE, but not SusF, prevented the growth of cells expressing SusD* (**Figure 2.3C**). This phenotype is not the result of a polar mutation from the *susE* deletion as immunoblots of whole cells with anti-SusF antibodies demonstrate the expression of *susF* (**Figure 2.3D**). These data suggest a unique role for SusE during growth on maltooligosaccharides.

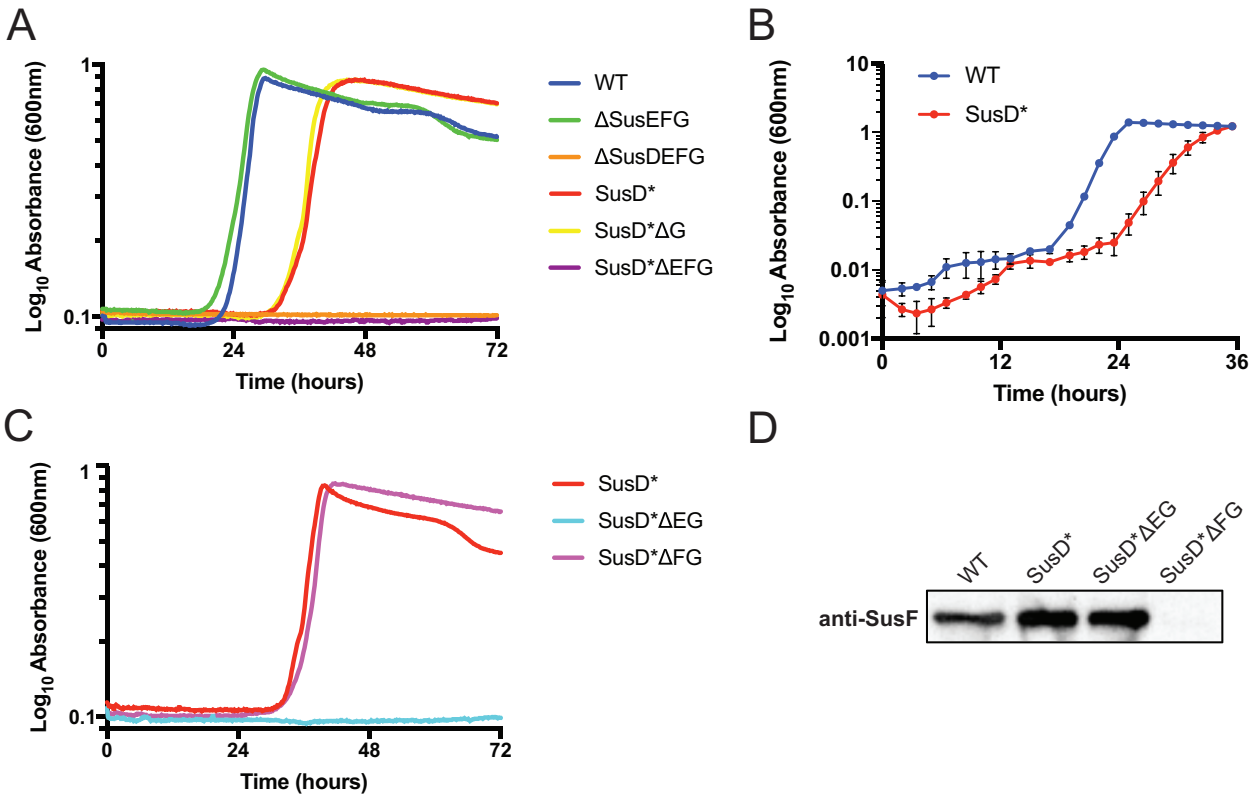

Figure 2.3. SusE compensates for the loss of the SusD starch binding site during growth on maltoheptaose

(A,C) Average growth curves of *B. thetaiotaomicron* strains expressing SusD* in different *sus* mutant backgrounds cultured on 2.5 mg/ml maltoheptaose. Identical growth experiments in glucose are shown in Fig S1ab. (B) Average growth curves of wild-type (WT) and SusD* *B. thetaiotaomicron* cultured on 2.5 mg/ml maltoheptaose with absorbance measured manually in a spectrophotometer. Biological triplicate cultures were cultured in glucose then back-diluted 1:800 into minimal media with maltoheptaose. (D) Western blot using anti-SusF serum against whole cell lysates from WT, SusD*, SusD*ΔEG, and SusD*ΔFG cultures. Strains were cultured in minimal media containing 5 mg/ml maltose and were arrested in logarithmic phase then normalized by $O.D._{600}$ for loading in SDS-PAGE.

SusE and SusF share multiple structurally homologous but functionally divergent domains

That SusE and SusF display divergent functions within Sus is noteworthy as SusE and SusF contain conserved structural characteristics suggesting redundant roles during starch catabolism. Both proteins have multimodular structures comprised of β-sheet rich starch-binding domains in tandem with an N-terminal Ig-like fold domain [74]. Proline residues are present between sequential domains of both proteins that presumably limit conformational flexibility, with the exception that SusE contains a putative flexible linker between the N-terminal domain and its first starch-binding domain, the Eb domain (see **Figure 2.4B** for schematic). SusF has three starch-binding domains, Fa, Fb, and Fc and SusE has two, Eb and Ec, named for their homology to the domains of SusF (**Figure 2.4B**). Among all five domains, Ec has the highest binding affinity for maltoheptaose and starch [74]. The Ig-like fold of SusE (E_N) was predicted to be similar to that of SusF (F_N).

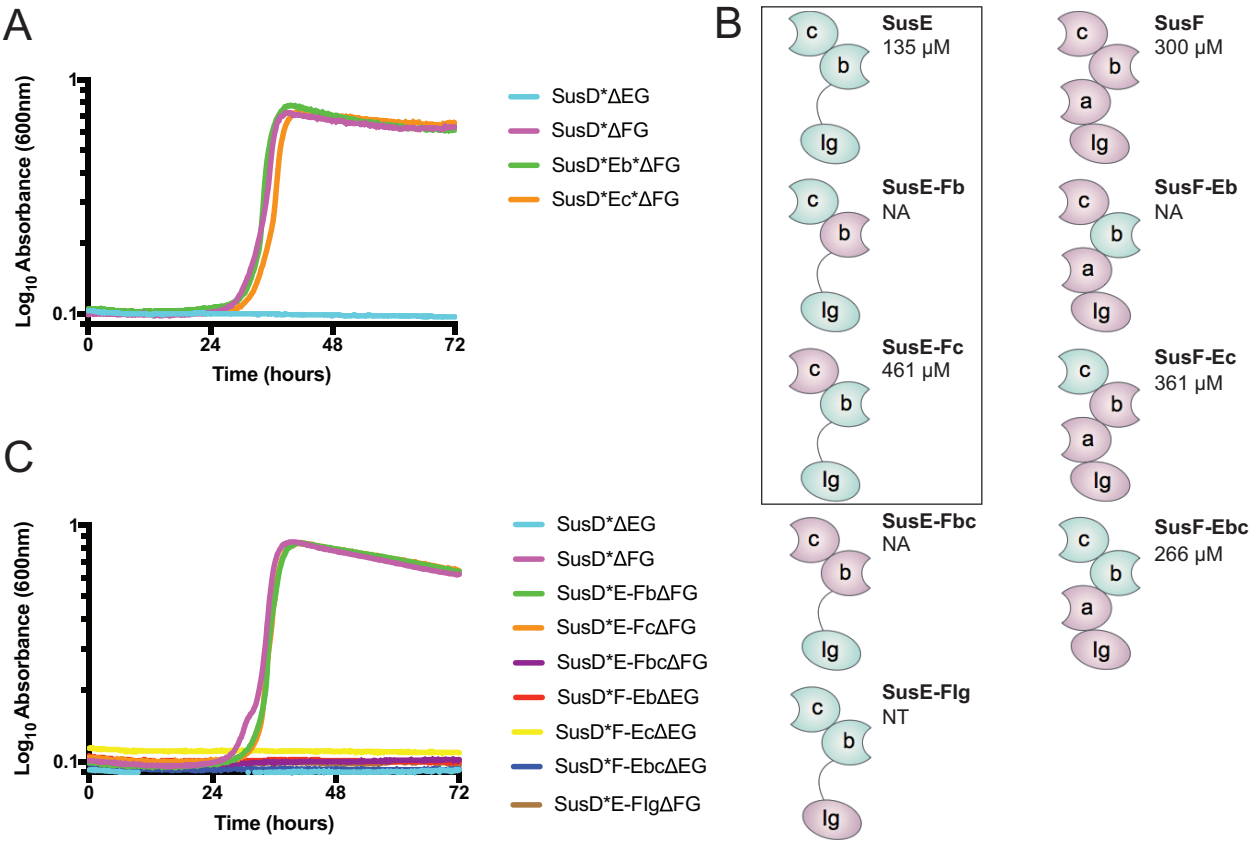

Figure 2.4. SusE and SusF are composed of functionally unique, but structurally homologous starch-binding domains

(A) Average growth curves of *B. thetaiotaomicron* strains expressing single starch-binding site mutants of SusE on 2.5 mg/ml maltoheptaose. Identical growth experiments in glucose are shown in Fig S1c. (B) Schematic of chimeric SusE and SusF domain structures. The K_d (mM) of each chimera for maltoheptaose as measured by isothermal titration calorimetry is reported. The K_d (mM) of the recombinant wild-type proteins is displayed for reference, as previously reported [74]. A box is placed around constructs that support growth in *B. thetaiotaomicron* expressing SusD*. NA denotes chimeric constructs that did not express in *E. coli*. NT denotes chimeric constructs not tested for recombinant expression. (C) Average growth curves of *B. thetaiotaomicron* strains expressing chimeric SusE and SusF on 2.5 mg/ml maltoheptaose. The mutants that displayed growth were SusD*ΔFG (light purple), SusD*E-FbΔFG (green), and SusD*E-FcΔFG (orange). Identical growth experiments in glucose are shown in **Figure 2.2D**.

To identify the domains of SusE that promote growth on maltoheptaose in the SusD* strain, we utilized mutant alleles of *susE* that we have previously reported that contain only one viable starch-binding site [74]. The SusEb* and SusEc* expressing strains, in which the Eb and Ec domain respectively cannot bind starch, were exchanged into the *B. thetaiotaomicron* chromosome in place of *susE*. Expression of these single-binding mutants with the SusD* allele revealed that both SusD*Eb* or SusD*Ec* strains grow on maltoheptaose with similar kinetics, demonstrating that both of the SusE starch-binding sites can compensate for the SusD* mutation (Fig. 3a). This was surprising as the K_d of Ec for maltoheptaose is ~ 20 μM and 50-fold higher than both SusEb and SusD [74]. Despite a K_d of 1 mM for maltoheptaose, native SusD efficiently supports growth on maltoheptaose emphasizing that high affinity binding to maltoheptaose is not required for transport through SusC (**Figure 2.3A**).

Given that either starch-binding site of SusE is adequate for growth on maltoheptaose, we sought to determine which domains of SusE must be present for growth. Exploiting the structural homology of SusE and SusF, we created a set of chimeric proteins in which Eb and Ec or Fb and Fc were swapped individually or in combination between SusE and SusF (**Figure 2.4B**). Immunofluorescent microscopy of fixed whole cells verified that these chimeric proteins were expressed by *B. thetaiotaomicron* and trafficked to the outer membrane (**Figure 2.5**); we have demonstrated that cells expressing site-directed mutants of SusE and SusF ($SusE^{C21A}$ and $SusF^{C20A}$) that retain the proteins in the periplasm cannot be labeled [74]. Recombinant expression and purification of these chimeras was performed to verify starch binding, however three of the seven mutants, SusE-Fb, SusE-Fbc, and SusF-Eb,

were not soluble when expressed in *E. coli* (**Figure 2.4C**). Since all of these proteins expressed and correctly trafficked in *B. thetaiotaomicron* we speculate that there may be some aspects of protein folding that are unique to *B. thetaiotaomicron*. Proteins that could be expressed in *E. coli* were assayed for maltoheptaose binding by isothermal titration calorimetry, and displayed K_d values similar to the wild-type proteins (**Figure 2.4B**, **Figure 2.6**) [74].

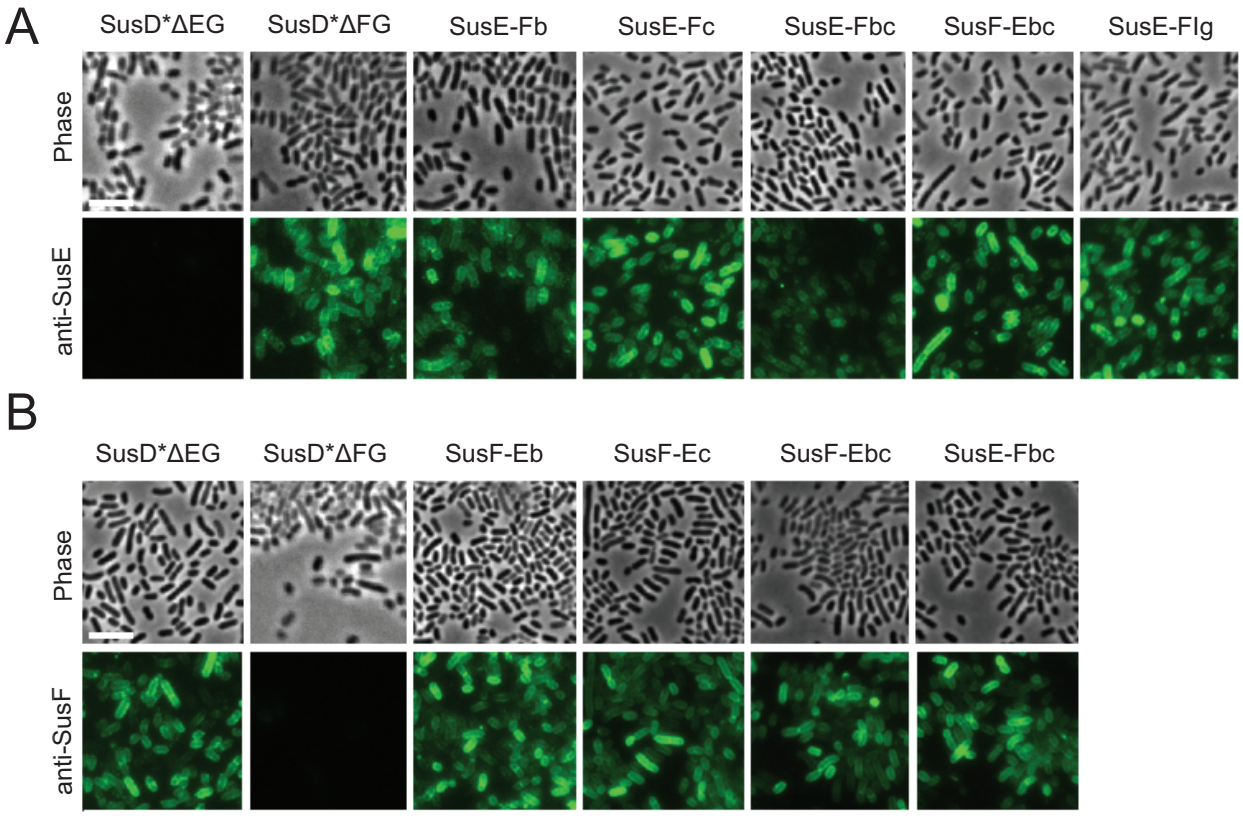

Figure 2.5. Surface staining of SusE and SusF chimeras

(A,B) *B. thetaiotaomicron* cells expressing a chimeric SusE or SusF in place of the wild-type alleles were visualized by immunofluorescence. Strains were grown in 5 mg/ml maltose and until mid-logarithmic phase. Non-permeabilized, formalin-fixed cells were probed with custom anti-SusE (A) or anti-SusF (B) serum then stained with anti-rabbit IgG conjugated to Alexa Fluor 488. Top and bottom panels show phase and fluorescence images, respectively. Scale bar = 5 μm.

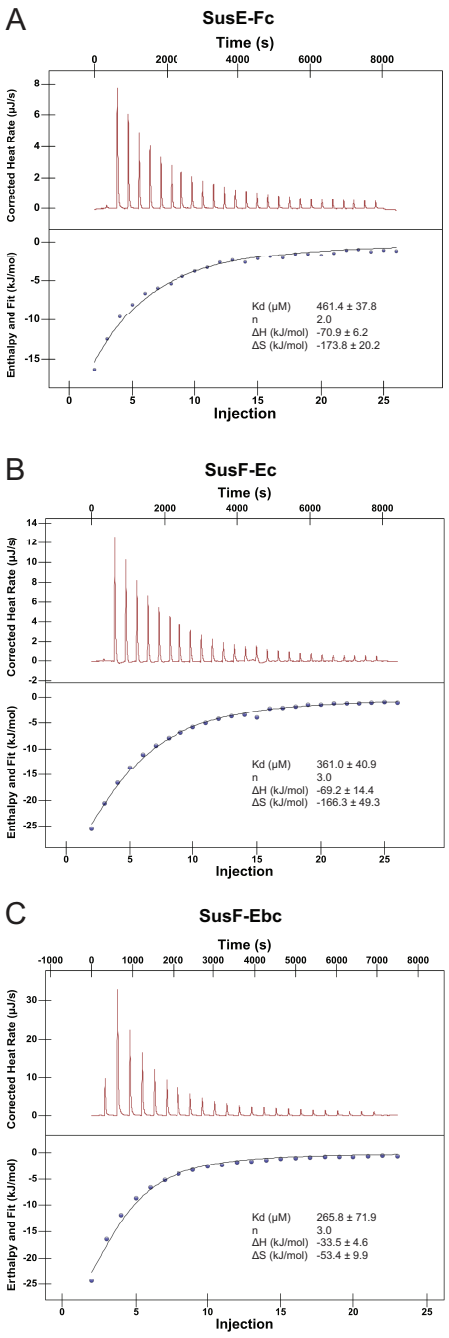

Figure 2.6. Isothermal titration calorimetry (ITC) of SusE and SusF chimeras with maltoheptaose

(A-C) All titrations were performed with buffer-matched protein and ligand samples (in 20 mM Hepes, 100 mM NaCl pH 7.0) and at 25 °C. In each case, the upper graph shows the raw heat signal for 23 x 1.0 µl injections of maltoheptaose into protein; the bottom graph shows the integrated heats with fits to an independent binding model, with N fixed to the number of binding sites in the protein.

When grown on maltoheptaose, chimeric SusE containing only one starch-binding SusF domain (SusD*E-FbΔFG and SusD*E-FcΔFG) was able to support growth as was the wild-type SusE allele in *B. thetaiotaomicron* expressing SusD* (**Figure 2.4C**). If both domains Eb and Ec were replaced with Fbc (SusD*E-FbcΔFG), then *B. thetaiotaomicron* could not grow, indicating that at least one of the SusE starch-binding domains is required. While the SusE-Fbc mutant could not be isolated from *E. coli*, the SusE-Fb allele that supports growth also could not be obtained recombinantly, supporting that lack of soluble expression in *E. coli* does not necessarily mean a lack of functional or correctly folded protein in *B. thetaiotaomicron*. None of the SusF chimeras possessing Eb or Ec domains were able to rescue growth in the SusD* expressing mutant. These observations suggest that the SusE starch-binding domains are necessary, but also that its N-terminal domain may be important. Indeed, our SusE mutant allele expressing the SusF N-terminal Ig-like domain (SusD*E-FIgΔFG), did not support growth on maltoheptaose although this chimera is expressed on the surface of the cell (**Figure 2.5**). Taken together, the ability of SusE to compensate for SusD* is dependent on its distinctive domains at both the N and C-terminus. However, we cannot rule out that the putative flexibility of SusE is also a required facet of its function. Note that the chimeric SusE proteins were created to retain the putative linker between the N-terminal domain and the following Eb or Fb domain.

Maltoheptaose binding by SusE is not required to promote growth

We have recently reported that immunoprecipitation of SusD results in co-isolation of both SusE and SusC, so we hypothesized that SusE provides a starch-

binding function within the proximity of SusC to guide glycans into the transporter [97]. Previous work provided evidence for an interaction between SusE and SusF, so we reasoned that SusF might contribute to glycan import via an interaction with SusE proximal to the SusCD complex [27]. However, the SusD*E*G (expressing wild-type SusF), SusD*E*ΔFG, and SusD*E*F*ΔG strains grow on maltoheptaose with an extended lag phase compared to the SusD*ΔFG that expresses a wild-type SusE (**Figure 2.7A**). We tested whether the additional lag from the SusD*E* strain could be rescued by a SusF chimera that has one or more domains of SusE. Interestingly, the only chimeric SusF that had any effect on growth of the SusD*E* strain was SusF-Ec, which abolished growth (SusD*E*F-EcΔG, **Figure 2.8**). Yet, SusF-Ec does not prevent growth when co-expressed with wild-type SusE (SusD*F-EcΔG, **Figure 2.8, 2.2J**). We speculate that the Ec domain on SusF-Ec is interfering with the Ec domain on SusE*, preventing its function in maltoheptaose transport. We also tested if the starch-binding sites on SusG enhance growth on maltoheptaose in the presence of SusD*E*F* as our previous work suggests that SusG dynamically interacts with the Sus complex [79]. However, a catalytically dead allele of SusG ($SusG^{D498N}$) does not improve growth on maltoheptaose (**Figure 2.7A**).

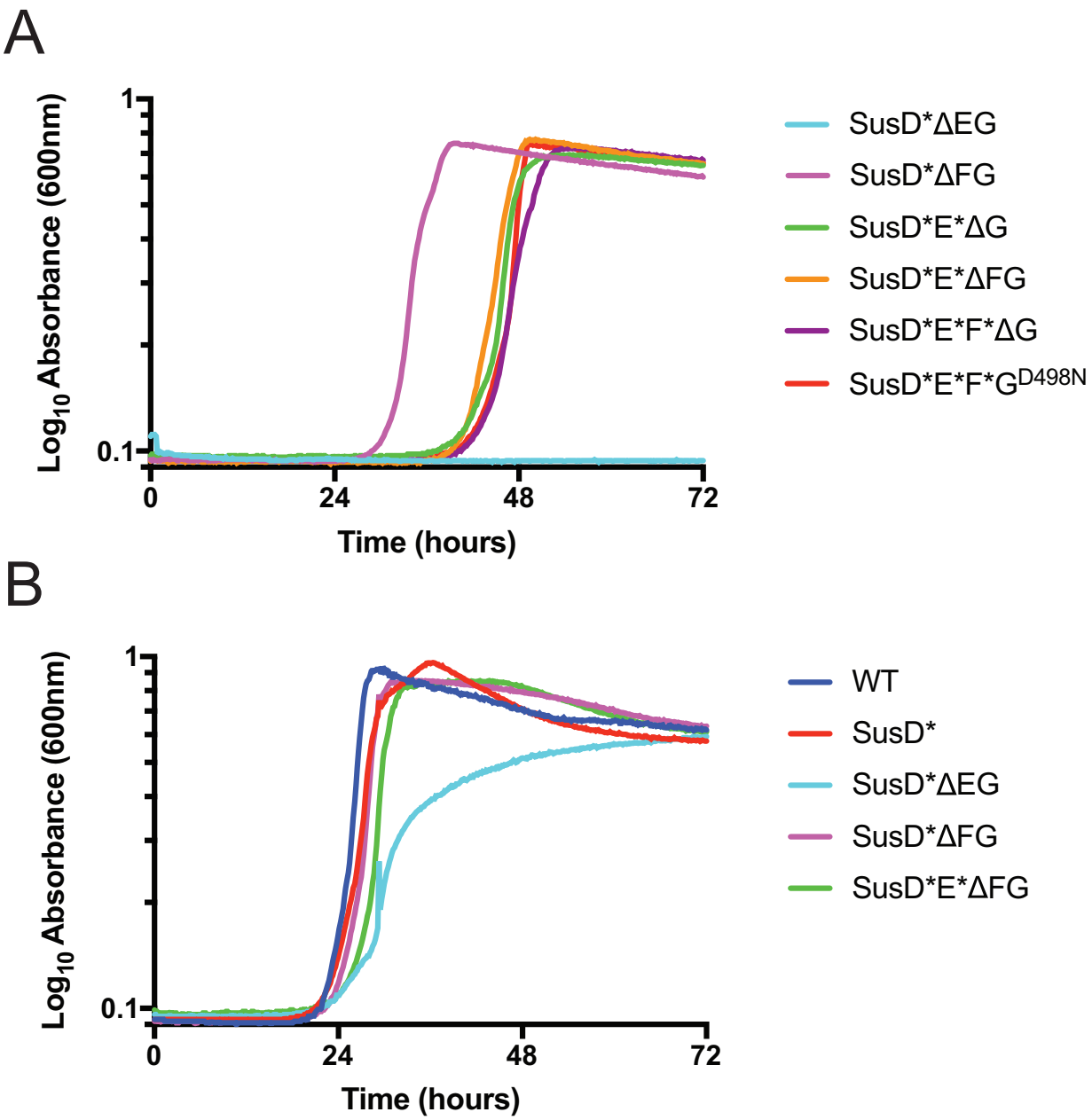

Figure 2.7. SusE does not require starch-binding sites for its unique functionality

(A) Average growth curves of *B. thetaiotaomicron* strains expressing starch-binding deficient SusD*E* in various *sus* mutant backgrounds on 2.5 mg/ml maltoheptaose or (B) 2.5 mg/ml maltoheptaose supplemented with 0.5 mg/ml maltose. Identical growth experiments in glucose are shown in **Figure 2.2E,F**.

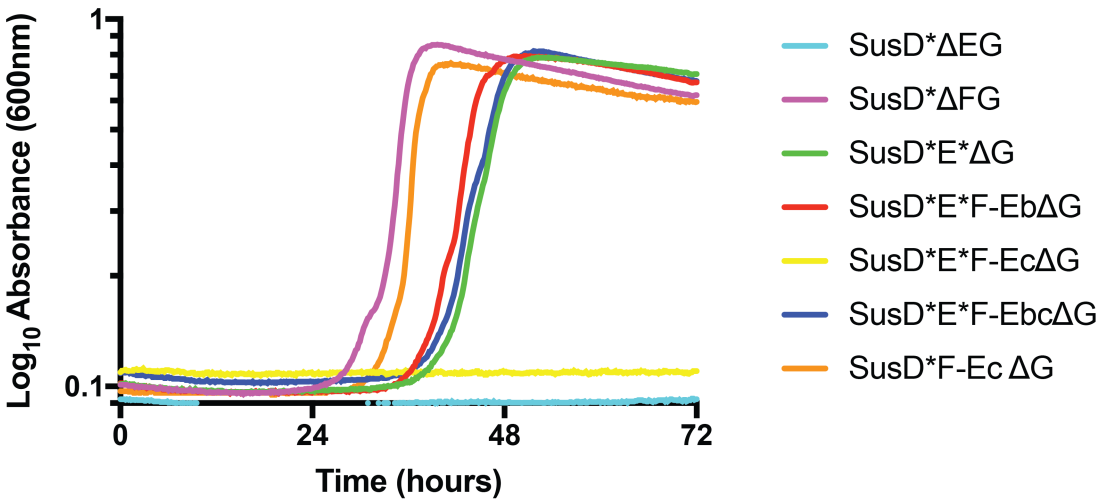

Figure 2.8. Chimeric SusF-Ebc Inhibits growth in *B. thetaiotaomicron* expressing SusD*E*

Average growth curves of *B. thetaiotaomicron* strains expressing starch-binding deficient SusD*E* combined with chimeric SusF on 2.5 mg/ml maltoheptaose.

The major observable difference in growth between our mutant strains is in the apparent lag, which occurs when these strains are cultured on maltoheptaose, not glucose (**Figure 2.2**). We attribute this lag to a defect in starch "sensing", or rather how quickly *sus* transcription is activated to support growth on starch; we have demonstrated previously that SusD* expressing cells require 100-1000x more sugar to maximally induce sus compared to wild-type cells [69]. However the addition of 0.5 mg/ml maltose, which does not require Sus for import and alone supports minimal growth of the strains (**Figure 2.9**), can upregulate *sus* and alleviate this lag [69]. To examine how sensing might be affecting our mutants, we cultured several SusD* mutant strains in minimal media with 5 mg/ml maltose, then subcultured into 2.5 mg/ml maltoheptaose with 0.5 mg/ml maltose. By inducing the *sus* operon with maltose, all strains grew comparably to wild-type with the exception of SusD*ΔEG, which now displayed growth on maltoheptaose but had a kinetic defect and lower maximum culture density (**Figure 2.7B**). Thus, it is likely that in the absence of maltose, a SusD*ΔEG strain cannot efficiently import and accumulate sugar to sufficient levels required to activate transcription and increase Sus protein levels on the cell surface; this type of kinetic coordination between the glycan levels for transcriptional activation and the activity of the periplasmic enzymes has been reported for the chondroitin sulfate and heparin targeting Sus-like systems in *B. thetaiotaomicron* [70]. However, the growth defect of the SusD*ΔEG strain on maltoheptaose plus maltose suggests that even when *sus* is activated, the presence of SusE with SusD* is required for effective maltoheptaose import. Taken together, these data support that SusE, but not SusF or SusG, displays

unique functionality during growth on maltooligosaccharides, and that its ability to support glycan uptake is not entirely dependent on its ability to bind sugar.

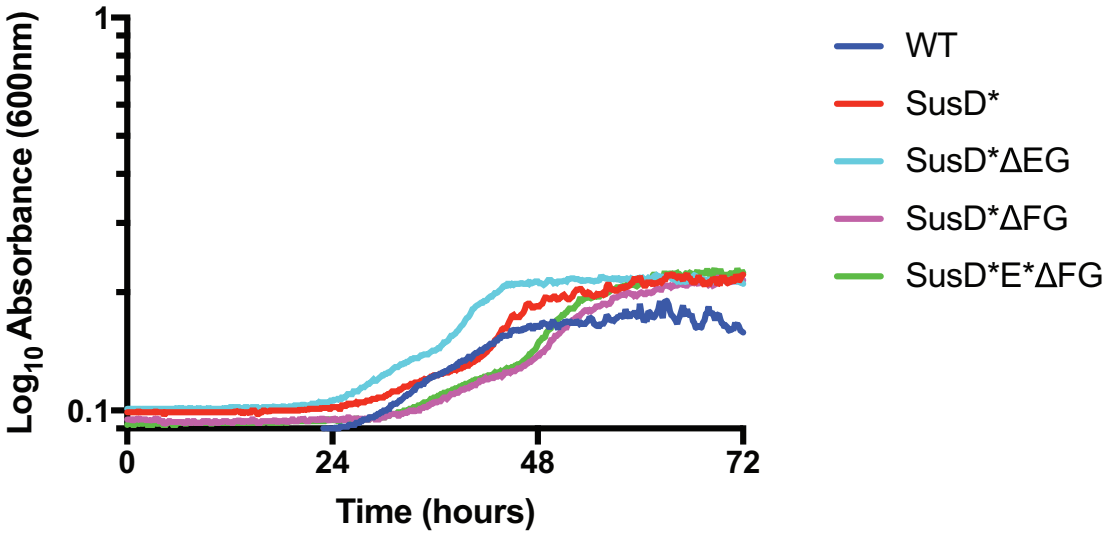

Figure 2.9. Growth of *B. thetaiotaomicron* on 0.5mg/ml Maltose

Average growth curves of WT and *B. thetaiotaomicron* strains expressing starch-binding deficient alleles on 0.5 mg/ml maltose.

Substrate binding by SusE provides a competitive advantage during growth on starch

Although maltoheptaose is a useful substrate to study the mechanism of maltooligosaccharide uptake, we wanted to know if SusE is similarly important for utilization of a starch polysaccharide like amylopectin. SusG is required for growth on starch, but in-frame deletions of *susEF* affect SusG transcription [41, 69]. Therefore we used the periplasmically localized mutants $SusE^{C21A}$ and $SusF^{C20A}$ [69] to test growth on starch in the absence of cell surface SusE and SusF. Cells were grown on maltose, then back-diluted into 5 mg/ml maize amylopectin with 0.5 mg/ml maltose to assess growth phenotypes. *B. thetaiotaomicron* expressing SusD grows on starch without SusE or SusF on the cell surface [69] (**Figure 2.10A**, $SusE^{C21A}$ and $SusF^{C20A}$). Cells expressing SusD* alone or in combination with $SusE^{C21A}$, $SusF^{C20A}$, or SusE* grow on starch but display a biphasic growth pattern with a more severe defect in the first phase. To assess the differences in growth rates, we quantified specific growth rates for each strain on glucose and starch during early ($O.D._{600}$ = 0.35) and late ($O.D._{600}$ = 0.75) exponential growth (**Figure 2.11**). Across the strains, growth on glucose was within standard error at both time points. On starch, cells expressing SusD were slightly faster in the first phase though not significantly so from the SusD* expressing strains (**Figure 2.11**). However, during late exponential growth on starch there was a significant decrease in growth rates in strains that expressed SusD* compared to those that expressed wild-type SusD, though there was no statistically significant difference among the SusD* expressing strains (**Figure 2.10B**). These observations support the idea that starch-binding by SusD is important for starch import, especially as starch may

become limited in later time points. That the presence of SusE was not a requirement for growth was surprising, but SusG may be compensating for this, either via its additional starch-binding sites or by generating small oligosaccharides such as maltose and glucose [56].

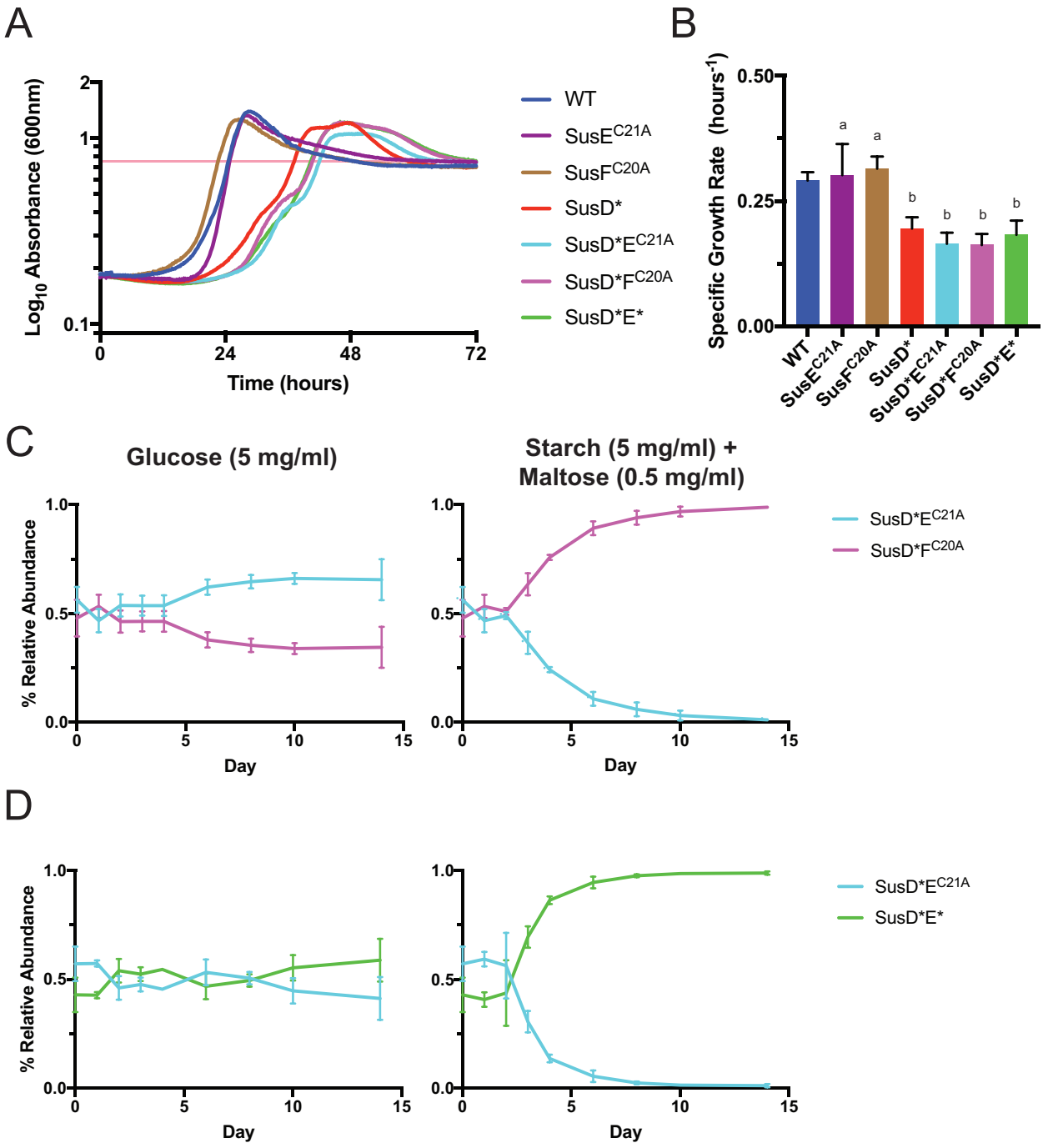

Figure 2.10. SusE provides a competitive advantage during drowth on starch

(A) Average growth curves of *B. thetaiotaomicron* strains expressing periplasmically localized SusE (C21A) and SusF (C20A) in combination with SusD on 5 mg/ml maize amylopectin supplemented with 0.5 mg/ml maltose. Identical growth experiments in glucose are shown in Fig S1g. The horizontal pink line indicates an $O.D._{600}$ of 0.75 at which specific growth rates were calculated for all strains. (B) Specific growth rates for all strains (n=3) grown in panel (A). Bars denoted with the letter “a” are not statistically significantly different from each other or WT. Bars denoted with the letter “b” are not significantly different from each other but are different from

WT with a P value < 0.05. Statistically significant differences were determined using the two-tailed unpaired Student's *t* test. (C,D) *In vitro* competitions of barcoded SusD* mutants in 5 mg/ml glucose and 5 mg/ml maize amylopectin with 0.5 mg/ml maltose. Relative abundance was calculated as the percent composition of a strain's DNA relative to the total DNA in the sample.

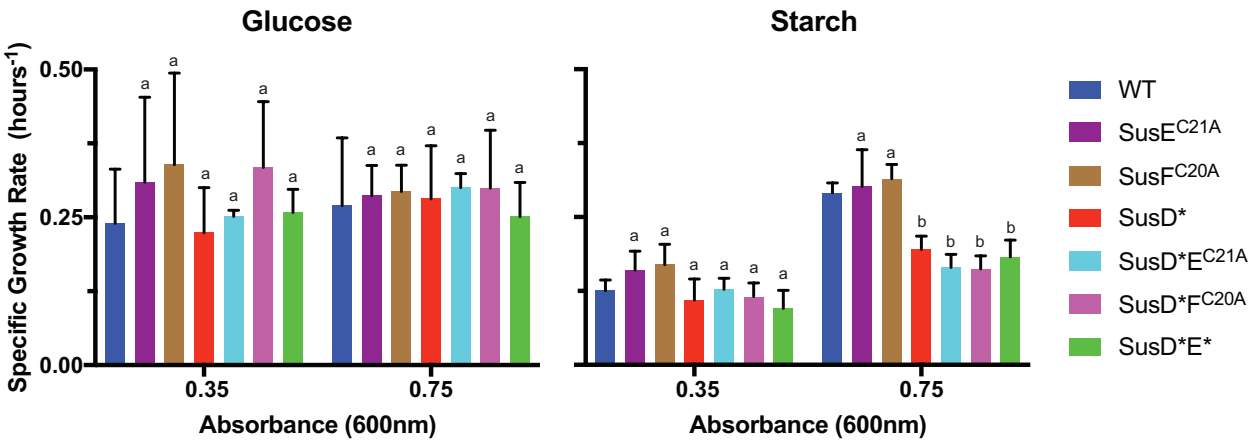

Figure 2.11. Specific growth rates for periplasmically localized SusE/F strains on 5 mg/ml starch with 0.5 mg/ml maltose

Specific growth rates are calculated at $O.D._{600}$ 0.35 and 0.75 by calculating the change in $O.D._{600}$/40 minutes/$O.D._{600}$ of the sample. Bars denoted with the letter "a" are not statistically significantly different from each other or WT. Bars denoted with the letter "b" are not significantly different from each other but are different from WT with a P value < 0.05. Statistically significant differences were determined using the two-tailed unpaired Student's *t* test.

Although the growth phenotypes displayed among the SusD* variants were similar, we wanted to know if the competitive fitness of *B. thetaiotaomicron* expressing SusD* was influenced by the presence of SusE, SusF, or SusE*. We performed *in vitro* competition experiments by co-culturing the SusD*E^{C21A} strain with either SusD*F^{C20A} or SusD*E*. These mutants were genetically tagged and passaged each day into minimal media containing glucose or starch and the relative abundance of each strain was quantified by qPCR (**Figure 2.10C,D**). The competitions in glucose may have displayed some stochastic changes in the abundance over time since it is unlikely the SusD*E^{C21A} mutant is better suited for growth on glucose than SusD*F^{20A}; this seems likely given the larger experimental error on the specific growth rates on glucose (**Figure 2.11**). Nonetheless, growth on starch resulted in a drastic decrease in the abundance of SusD*E^{C21A}, as it was outcompeted by both SusD*F^{C20A} and SusD*E* within the same two week time frame. These results not only underscore that SusE does not need to bind starch to support growth, but also raise the possibility that SusE provides a fitness advantage by facilitating glycan uptake from the environment.

Surface starch-binding proteins coordinate oligosaccharide uptake based on their length

In the Bacteroidetes, oligosaccharides are generated from carbohydrate-active enzymes anchored to the cell surface or secreted via outermembrane vesicles into the environment [98-100]. Our data suggest that surface glycan-binding proteins like SusD, SusE, and SusF are important for optimal glycan uptake through SusC, and we hypothesized that these starch-binding proteins could drive the preferential uptake of

certain lengths of maltooligosaccharides. To test this we grew *B. thetaiotaomicron* on a mixture of maltooligosaccharides spanning the degrees of polymerization (DP) 1 – 40 glucose units (a commercial preparation marked as DP10 – 40, but HPAEC-PAD analysis of the starting mixture revealed the presence of smaller sugars ranging from DP1 – 10) (**Figure 2.12A, Figure 2.13**). All HPAEC-PAD chromatograms reported in this study are compiled within **Figure 2.13**. Glucose, maltose, and maltotriose all support the growth of *B. thetaiotaomicron* independent of Sus [24], but the majority of glycans in the mixture are > DP7 (**Figure 2.12A**). Longer sugars in the range of DP32 – 40 fell below our detection limit here, but were detected in subsequent experiments. We considered that some of the longer oligosaccharides (> DP7) might not be transported efficiently, or at all, in the absence of SusG. Consistent with this notion, we observed that ΔSusG grows to a noticeably lower maximum O.D. compared to wild-type (**Figure 2.14A**). Furthermore, we found this trend to be consistent across all other strains that lack SusG (**Figure 2.14B**). Growth on DP10 – 40 by SusD*, SusD*E^{C21A}, and SusD*F^{C20A} all displayed similar kinetics, likely due to the enzymatic activity of SusG. Conversely, SusD*ΔG, SusD*ΔFG, and SusD*E*ΔG grew similarly on the DP10 – 40 mixture but the combination of starch-binding deficient mutations and loss of SusG resulted in these strains exhibiting an even lower maximum O.D. Consistent with a unique and important role in oligosaccharide uptake, the growth of SusD*ΔEG had a notably longer lag and the lowest maximum O.D. among these strains, underscoring the importance of SusE during starch uptake. The growth of SusD*ΔEG on DP10 – 40 is likely supported by the presence of maltose which induces *sus*.

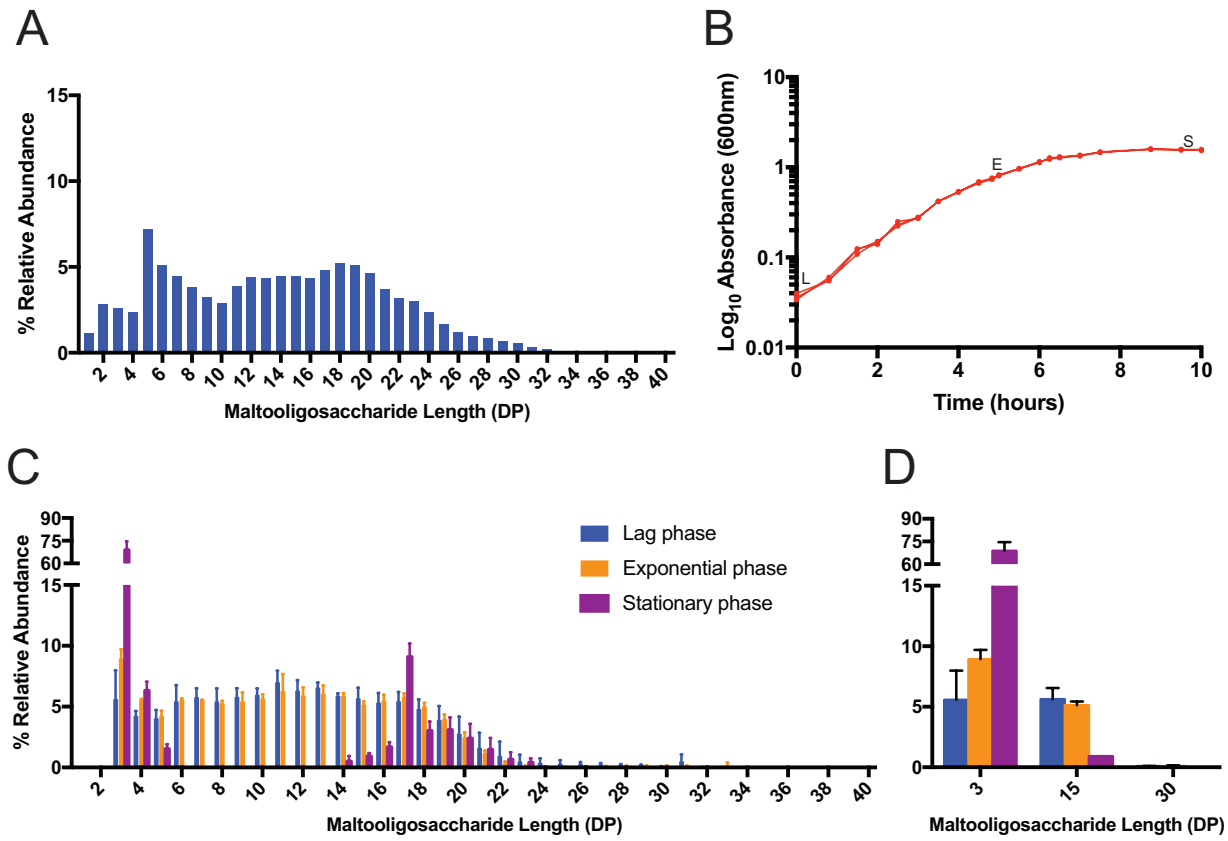

Figure 2.12. Maltooligosaccharides profile cell-free culture supernatants during ΔSusG growth on DP10 – 40 maltooligosaccharide mix

(A) Relative abundance of maltooligosaccharides in the DP10 – 40 maltooligosaccharide mixture as measured by HPAEC-PAD (n=1). (B) Growth curves (n=3) of ΔSusG on 5mg/ml DP10 – 40. L, E, and S denote when the cell-free supernatant samples for Lag, Exponential, and Stationary phases were taken. (C) Relative abundance of maltooligosaccharides from ΔSusG cell-free culture supernatants at lag, exponential, and stationary phase growth (n=3). Mid-range and longer oligosaccharides fell below our limit of detection in the stationary phase samples. Maltooligosaccharides were purified from media components resulting in the loss of glucose and maltose. (D) Enlargement of small (DP3), medium (DP15) and large (DP30) maltooligosaccharides in cell-free culture supernatant of ΔSusG cultures. Error bars indicate standard deviations across three biological replicates.

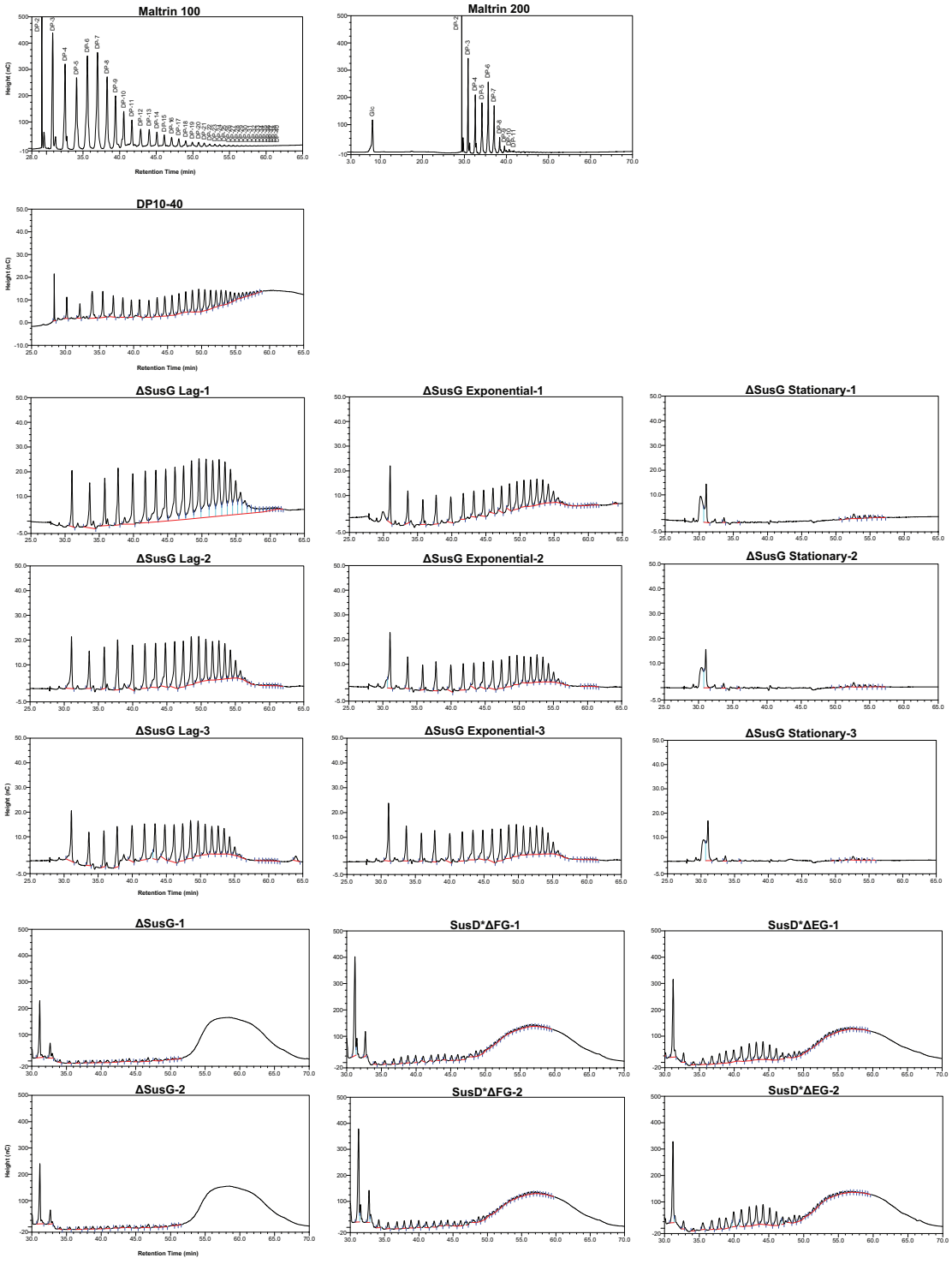

Figure 2.13. Compiled HPAEC-PAD chromatogram traces

HPAEC-PAD chromatograms for each experimental replicate in this study. Representative Maltrin 100 and 200 standards that were used to determine oligosaccharide retention times are shown here with annotated peaks. Peak intensity as measured by pulsed amperometric detection (PAD) is displayed as nano Coulombs (nC).

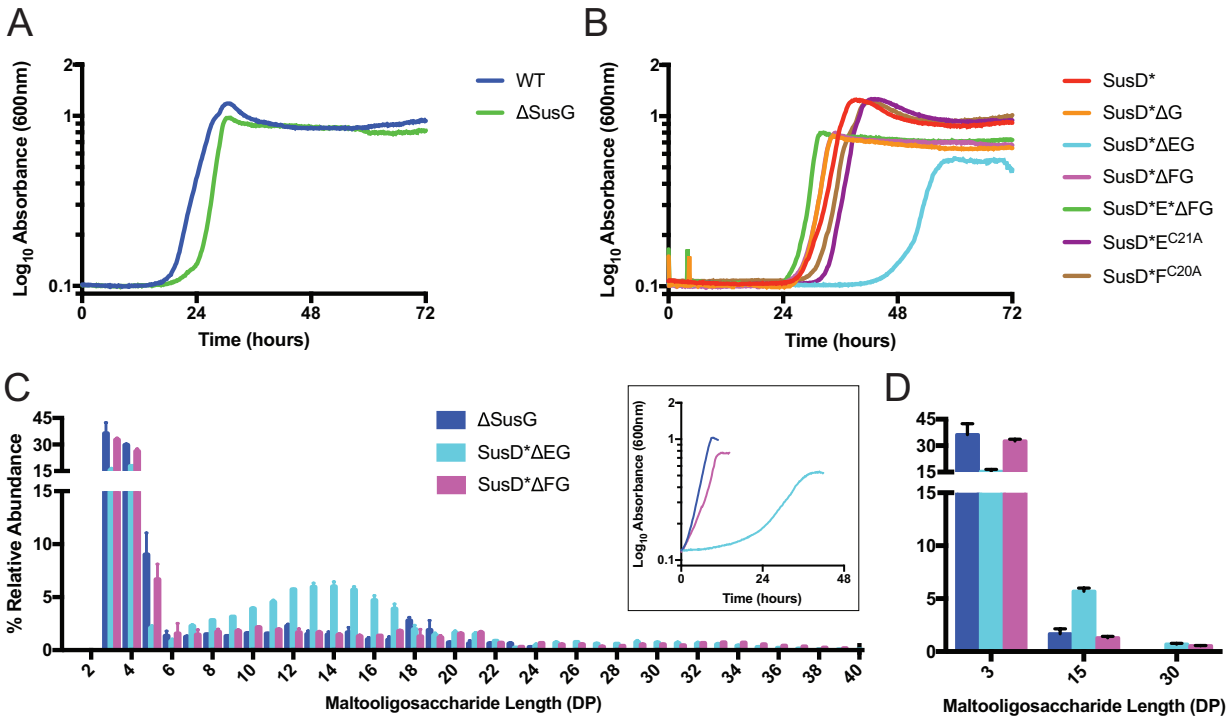

Figure 2.14. Starch-binding proteins facilitate the uptake of maltooligosaccharides in a size dependent manner

(A) Average growth curves of *B. thetaiotaomicron* WT and DSusG strains on 5 mg/ml DP10 – 40 maltooligosaccharide mix. Identical growth experiments in glucose are shown in **Figure 2.2H**. (B) Average growth curves of *B. thetaiotaomicron* SusD* expressing mutants on 5mg/ml DP10 – 40. Identical growth experiments in glucose are shown in **Figure 2.2I** (C) Relative abundance of DP3 – DP40 maltooligosaccharides in cell-free culture supernatants at stationary phase (n=2). Maltooligosaccharides were purified from media components resulting in the loss of glucose and maltose. (D) Enlargement of representative small (DP3), medium (DP15) and large (DP30) maltooligosaccharides in cell-free culture supernatant of stationary phase cultures. Error bars indicate standard deviations across two biological replicates.

To examine which maltooligosaccharides are depleted over time during growth, we grew ΔSusG on DP10 – 40 and measured glycan content in cell free supernatents from lag, exponential, and stationary phase (**Figure 2.12B**). Note that the preparation of media samples for HPAEC-PAD analysis resulted in the loss of glucose and maltose, and so these glycans are omitted from the analysis. Lag and exponential phase ΔSusG cultures had a similar distribution of oligosaccharides with a significant increase in the relative abundance of DP3 and DP4 in stationary phase (**Figure 2.12C**). An increase in a sugar's relative abundance within the sample indicates that it is either being generated from the break down of a larger oligosaccharide, or it is being imported less frequently compared to other sugars, or both. Because we can detect glycolytic activity in ΔSusG culture supernatants due to the presence of intracellular or secreted enzymes within culture supernatants, we attribute the increase of DP3 and DP4 abundance at least partially to enzymatic activity (**Figure 2.15**). Interestingly, there was a dramatic decrease in DP5 – DP16 but an increase in DP17 content and steady levels of larger DP sugars from exponential to stationary phase. If unidentified glycolytic activity in the supernatant were wholly responsible for the loss of mid-range glycans we would expect a concomitant decrease in larger glycans as well, rather than a discrete increase in glycans of a particular DP length. Therefore we conclude from these data that the SusCDEF complex can import, and may select for, maltoolisaccharides of DP5 – 16.

Figure 2.15. Stationary phase ΔSusG supernatants display glycolytic activity

Thin layer chromatography indicating the presence of glycolytic enzymes from stationary phase ΔSusG cultures grown in minimal media containing maltose. Ladder contains 5 mmols of glucose, maltose, and maltotriose. Supernatant (2 µl) was collected from cell-free stationary phase ΔSusG cultures. DP10 – 40 lane contains 100 µg of DP10 – 40 from MM containing 5 mg/ml DP10 – 40. Supernatant+DP10 – 40 contains both the ΔSusG culture supernatant and MM-DP10 – 40. All samples were incubated at 37°C for 1 hour before being loaded on the TLC plate. Degradation products, primarily maltotriose, are detected in the final lane indicating the presence of glycolytic enzymes in culture supernatant.

We investigated the contributions of SusE and SusF to maltooligosaccharide preference by characterizing the portfolio of oligosaccharides in the media from parallel cultures of SusD*ΔFG, SusD*ΔEG and ΔSusG cells at stationary phase from growth on DP10 – 40. Once again, these strains displayed different growth kinetics and maximum culture densities, suggesting inefficient or selective uptake of certain glycans in the mixture (**Figure 2.14B,C** inset). Like the first ΔSusG growth experiment, there is an enrichment of DP3 – 4 compared to other sugars, which may be because these are less frequently taken up, and there is likely some glycolytic activity in the media. Interestingly, sugars > DP23 were not found in ΔSusG cultures but were detected in SusD*ΔFG and SusD*ΔEG cultures, which may in part explain why these strains grow to a lower maximum O.D. SusD*ΔEG, which displays the most severe growth defect on DP10 – 40, is less efficient at taking up sugars DP7 – 18 and removes more of the small sugars 3 – 5 compared to ΔSusG and SusD*ΔFG cells. These data suggest that SusE is needed to efficiently access mid-range and longer sugars, and in the absence of SusE the cells may need to scavenge smaller sugars to support growth (**Figure 2.14D**). Although the energetics of maltooligosaccharide import in the Bacteroides is largely unknown, the uptake of mid-length glycans through the Sus likely minimizes the number of transport events needed to support growth and may be a cost-effective strategy for glycan capture.

Discussion

Members of the gut microbiota compete for carbon and energy sources to survive in the densely populated colonic environment and the Bacteroidetes that

dominate this niche rely on sets of cell surface proteins to recognize, degrade, and import dietary polysaccharides. Glycan transport across the outermembrane is a critical feature of these systems that allows for the complete depolymerization of polysaccharides in the periplasm, which prevents the release of monosaccharides to neighboring species [101]. Hence, understanding how these bacteria import carbohydrates can provide knowledge for the strategic manipulation of select species in the gut. Here, we investigate glycan uptake via the prototypical Sus of *B. thetaiotaomicron*. Our findings have uncovered that the transporter SusC imports maltooligosaccharides in a manner that requires SusD and SusE, but is not contingent on their capacity for starch-binding, suggesting assembly of the complex is most important. Furthermore, these data suggest that the protein-protein interactions that dictate Sus complex assembly can tune how the cell acquires starch. These unexpected observations raise further questions about the underlying mechanisms of glycan transport by homologous Sus-like systems, particularly the relationship between SusC-like TBDTs and their cognate glycan-binding proteins. Two recent crystal structures of *B. thetaiotaomicron* SusCD-like complexes reveal that the SusD-like protein sits atop the SusC TonB-dependent transporter (TBDT), with the ligand-binding site directed into the opening of the barrel [83]. Molecular dynamics simulations of the complex for peptide import reveal that binding of the ligand by SusD protein and the internal plug of the TBDT stabilizes the closed complex, triggering the import event. In the absence of ligand, SusD can open to repeat this cycle. A difference between the two SusCD-like crystal structures was the presence of additional PUL-encoded lipoproteins that co-purified with the complex. The putative peptide-targeting SusCD-like complex included

proteins BT2261 – 2264, with two lipoproteins, BT2262 and BT2261. BT2261 and BT2262 are comprised of Ig-like folds, akin to SusE and SusF, and wrap around the back of the CD-like complex; these proteins seems to affect the movement of SusD in molecular dynamics simulations, suggesting they contribute to the open/close mechanism of the transporter. However, the precise role of BT2261 and BT2262 in selection of ligand or bacterial growth has not been elucidated.

Glycan-binding lipoproteins akin to SusE and SusF, sometimes referred to as "SusE-positioned genes" or surface glycan binding protein-B (SGBP-B), functionally differ from SusD and its homologs. The presence of SusD-like proteins is required in some Sus-like systems like that for the uptake of xyloglucan in *B. ovatus* [102], but is less important for the uptake of levan and not at all for the uptake of fructo-oligosaccharides in *B. thetaiotaomicron* [36]. SusEF-like SGBPs are encoded in PULs throughout the gut Bacteroidetes suggesting that they play a critical role in glycan capture [15, 39]. However, the importance and nature of their role may vary depending on the characteristics of the substrate. In the Sus, SusE appears to have evolved an intriguing role in maltooligosaccharide uptake (**Figure 2.16A**). Our data support that SusC and SusD interact to form a high affinity transporter that facilitates efficient growth on maltoheptaose. SusD*, in the absence of other starch-binding proteins, cannot support growth, possibly by stably interacting with SusC and preventing the introduction of sugar into the transporter. This sort of closed complex absent of substrate is seen in the SusCD-like crystal structure of BT1763/1762 [83]. SusE can uniquely resuscitate growth independent of its starch-binding sites, which we speculate is accomplished by SusE interacting with the core SusCD complex and allowing SusD to open for import.

Curiously, the starch-binding sites on SusF and SusG do not enhance growth kinetics in a SusD*E* strain. This observation highlights the idea that starch-binding on the cell surface is not sufficient to assist in transport, and that the interaction of SusE with the SusCD transport complex may give it a privileged proximity near the opening of the pore.

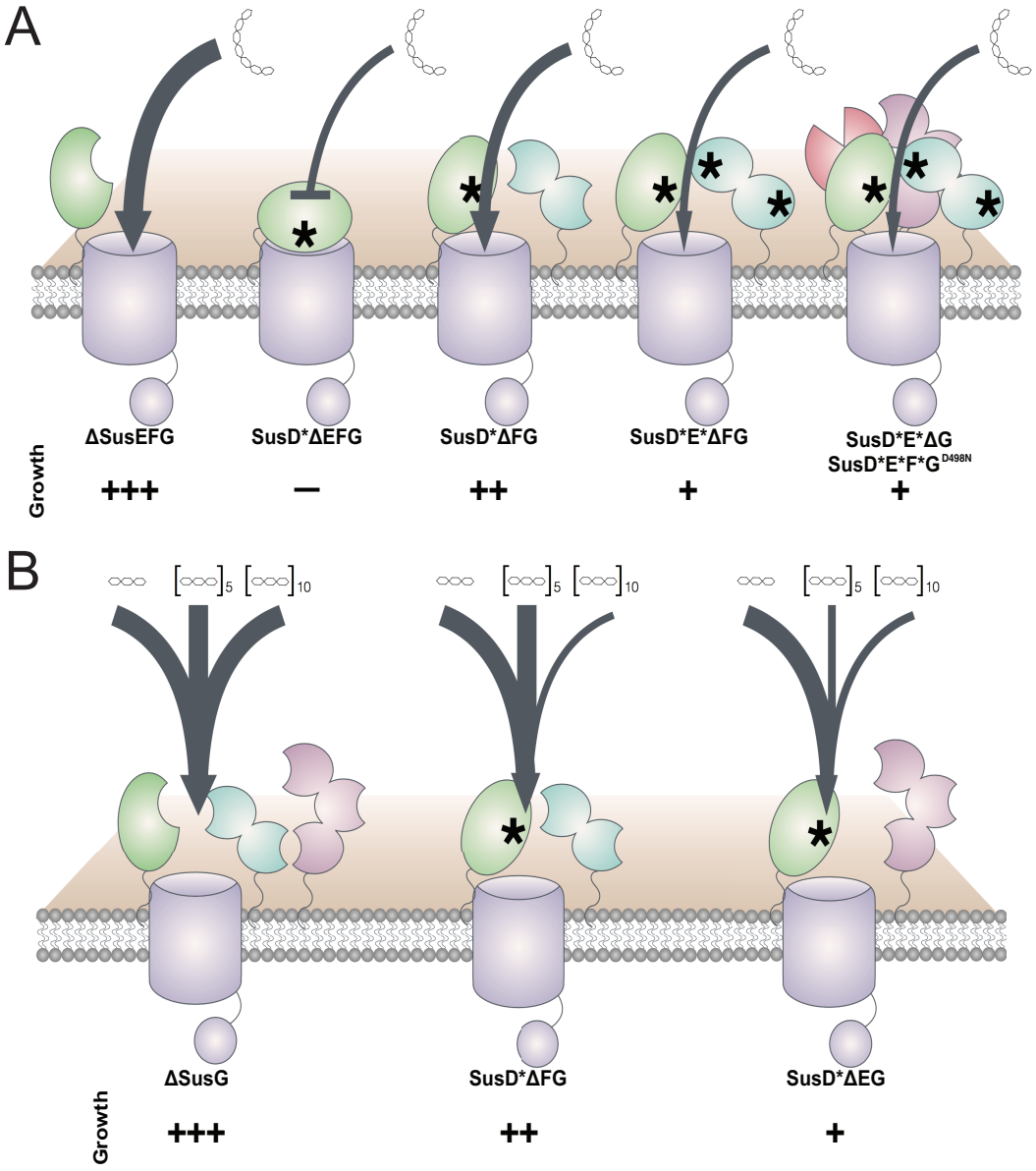

Figure 2.16. A model for maltooligosaccharide uptake facilitated by the SusEF starch-binding proteins

Asterisks indicate a site-directed mutant that no longer binds starch. The Sus proteins are colored: SusC, purple; SusD, green; SusE, blue; SusF, magenta, SusG, pink. (A) The relative growth of the cells on maltoheptaose is displayed with more + indicating less lag time and more efficient growth. SusEF are not necessary for *in vitro* growth on maltoheptaose, but the presence of SusE is required in a SusD* background. (B) The relative growth of the cells on maltoligosaccharides of various DP is displayed with more + indicating a greater maximum O.D. and better utilization of the glycans within the mixture. The thickness of the arrow indicates the ability of the cells to take maltooligosaccharides of DP3, DP15 and DP30. Cells expressing SusCDEF (ΔSusG) can utilize all sizes of maltooligosaccharides, while cells expressing only SusE or SusF in a SusD* background are impaired for the uptake of large and mid- to-large glycans respectively.

That bacteria would evolve glycan uptake systems that import larger oligosaccharides is logical from an energetics viewpoint – it saves the energy of active transport when larger oligosaccharides can be imported at the same relative energetic "cost" as smaller ones. Although a different type of import system from the Sus, ATP-binding cassette (ABC) transporters for maltodextrins in Gram-positive bacteria such as *Streptococcus pneumoniae* and *Eubacterium rectale* have been observed to specifically target the uptake of oligosaccharides larger than maltose [14, 103]. In these cases, the surface digestion of polysaccharide is coordinated with the specificity of the ABC transport solute-binding protein. For example, in *Streptococcus pneuomoniae* the surface enzyme SpuA digests glycogen into a ladder of fragments of up to at least 30 residues in length, and the MalX solute-binding protein aids in the transport of maltooligosaccharides of up to 12 residues, based upon the size of the glycans preferentially depleted from culture supernatants of cells grown on glycogen [103]. This is in contrast to *E. coli*, in which the maltose-binding protein MalE can apparently transport glycans up to eight or ten glucose residues, but growth defects are observed on maltooligosaccharides larger than maltoheptaose, due in part to the inefficiency of the uptake of these longer sugars [104, 105]. Despite differences in mechanism, these maltooligosaccharide ABC transporters and the Sus are similar in that they provide a strategy for the specific uptake of maltooligosaccharides, however the Sus may be unique in its ability to target even longer glycans.

Our data suggest that the collection of Sus proteins at the surface may dictate the length of maltoligosaccharides captured by *B. thetaiotaomicron*. Surprisingly, the ΔSusG strain is capable of importing even the longest range of oligosaccharides (DP25

– 40), as suggested from the lack of detection of these sugars in the culture supernatent, although they are detected in SusD*DEG and SusD*DFG cultures. Because these sugars are much larger than the substrates targeted by classically studied TonB-dependent transporters [84], it seems probable that SusC has been adapted to target larger substrates derived from polysaccharides. Indeed, that SusC-like transporters for other complex glycans such as a-mannan, rhamnogalacturonan-II, chondroitin sulfate, and heparin or heparan sulfate can import very large oligosaccharides has been suggested or observed [35, 70, 93, 96]. SusG is an endoamylase that can conceivably generate large oligosaacharides when hydrolyzing starch, yet *in vitro* digests yield glucose and maltose [56]. Perhaps at the cell surface, the product profile of SusG is altered to include longer sugars, or the Sus can also sequester long oligosaccharides generated from neighboring species. Our work has also shown that maltooligosaccharides of varying length can be imported with different efficiencies, and that starch-binding proteins can in part modulate these differences (**Figure 2.16B**). We show here that SusE and SusF can greatly effect the uptake efficiency of maltooligosaccharides at the cell surface. SusD*ΔEG displays a much lower capacity to efficiently transport mid- and long-range maltooligosaccharides of DP5 – 16. This forces the SusD*ΔEG strain to grow on smaller sugars that provide a poor return on the energetic investment to actively import those substrates. SusD*ΔFG has a similar but less critical deficiency during uptake as well. That these surface glycan-binding proteins can adapt the cell to access different lengths of the same type of glycan is currently unexplored, but it may provide a mechanism explaining how bacteria can partition into nutrient niches or how they can share nutrients while foraging on the

same glycans. The Bacteroidetes acquire large, complex, and heterogenous substrates for carbohydrate catabolism in order to remain competitive in the gut ecosystem. A mechanistic understanding of nutrient acquisition in the gut microbiome is a key prerequisite to intelligently manipulating this ecosystem to our benefit [106, 107].

Materials and Methods

Bacterial strains and culture conditions

For these experiments and to generate all of the mutant *sus* strains used in these experiments, the *B. thetaiotaomicron* VPI-5482 Δ*tdk* strain was employed to facilitate allelic exchange, as previously described [42, 74], and is the parent strain for all of the mutations within this this work. For clarity we refer to the Δtdk strain as wild-type, as this parent strain retains a wild-type *sus* locus. Mutations were generated using the counterselectable allelic exchange vector pExchange-*tdk* as previously described [42].. The *sus* alleles for all Sus* mutants are included in Table 1 and were validated in our previous studies, as referenced. A table of strains used in this study is provided in **Table 2.1**. SusE-F chimeric constructs were designed based upon the known structures of these proteins. Oligonucleotides used in this study are listed in **Table A.1**.

B. thetaiotaomicron was cultured in a 37°C Coy anaerobic chamber (5% H_2/10% CO_2/85% N_2) from freezer stocks into tryptone-yeast extract-glucose (TYG) medium [108] and grown for 24 hours, to an O.D. ~1.0. The following day cells were back-diluted 1:100 into Bacteroides minimal media (MM) including 5 mg ml^{-1} glucose or maltose (Sigma) as noted and grown overnight (16 hours). For kinetic growth experiments, MM-glucose or MM-maltose grown cells were back-diluted 1:200 into MM with the

experimental carbohydrate, and in parallel to MM with glucose. Thus both glucose controls and experimental starch and oligosccharide grown cultures were started at the same initial $O.D._{600}$ of 0.1 in the plate reader (or 0.02 in a 1cm path length spectrophotometer at $O.D._{600}$). All glucose controls for each set of growth experiments are displayed in Supplemental Figure 1. The substrates used for comparison to parallel glucose-grown cultures included: 2.5 mg ml^{-1} maltoheptaose (Carbosynth), 5 mg ml^{-1} maize amylopectin (Sigma), or 5 mg ml^{-1} DP10 – 40 (Elicityl). Kinetic growth experiments were performed at 37°C in 96 well plates and $O.D._{600}$ were recorded every 10 – 30 min. All plate reader growth experiments were performed in 3 – 10 replicates and the averages are reported in each figure here. However, all biological experiments were repeated at least twice to verify consistent growth phenotypes from day to day. Specific growth rates were calculated at $O.D._{600}$ 0.35 and 0.75 as the $\Delta O.D._{600}$ $\Delta time^{-1}$ $O.D._{600}^{-1}$. The change in O.D. was calculated over a 40 minute duration. Manual growth curves were performed in 16mm glass culture tubes and $O.D._{600}$ measurements were taken from the same tubes by a Genesys20 spectrophotometer. The spectrophotometer was blanked with media plus maltoheptaose including those curves for the starch media, which has a background absorbance of 0.2. The calculation of the specific growth rates for the starch grown cultures was performed after subtracting this background absorbance from the time points. *B. thetaiotaomicron* was cultured in culture tubes as previously specified except cultures were started from 1:800 back-diluted overnight grown cells in MM plus glucose, or maltose for the starch-grown cultures. Specific growth rates were calculated as the $\Delta O.D._{600}$ $\Delta time^{-1}$ $O.D._{600}^{-1}$. The change in O.D. was calculated over a 40 minute – 2 hour duration.

Cloning and recombinant protein expression

Chimeric SusE and SusF alleles were PCR amplified from genomic DNA for ligation-independent cloning into pETite N-His vector (Lucigen Madison, WI) according to the manufacturer's instructions. The N-terminal primers introduced a TEV-cleavable site between the mature protein that lacks a signal sequence and the His tag as well as a mutation C21A and C20A of SusE and SusF respectively to produce soluble TEV-cleavable His-tagged proteins. Chimeric SusEF-containing pETite plasmids were transformed into Rosetta (DE3) pLysS cells (EMD Biosciences) and plated on LB agar containing 50 μg ml^{-1} kanamycin (Kan) and 20 μg ml^{-1} chloramphenicol (Cm). After 16 hours of growth at 37°C, colonies from plates were used to inoculate 1 l of terrific broth plus Kan and Cm for growth at 37°C. Cultures were grown to an $O.D._{600}$ of ~0.6 before 0.5 mM IPTG was added and cells were grown at room temperature (~20°C) for an additional 20h. Cells were harvested by centrifugation and cell pellets were flash frozen in liquid N_2 until purification.

Recombinant protein purification

Cell pellets were resuspended in 50 mls of His Buffer (25 mM NaH_2PO_4, 500 mM NaCl, 20 mM imidazole, pH 7.4) and were lysed by sonication. Lysates were centrifuged at 20,000 x g to remove intact cells and the soluble lysate containing. His-tagged chimeric SusEF proteins were purified using a 5 mL Hi-Trap metal affinity cartridge (GE Healthcare) charged with $NiSO_4$ according to the manufacturer's instructions. Cell lysates were applied to the column in His Buffer and proteins were eluted with an

imidazole (20-300 mM) gradient. The His-tag was removed with a 2 hour incubation with recombinant TEV (1:100 molar ratio of TEV to protein) at room temperature, followed by an overnight incubation at 4°C while dialyzing into His buffer with 20 mM imidazole. Affinity purification using a Ni-charged Hi-Trap affinity cartridge was repeated to remove the His-tag, uncut protein and His-tagged TEV, while the cleaved protein was collected as the flow through. The protein was then dialyzed against a storage buffer (20 mM HEPES, 100 mM NaCl, pH 7.0) and concentrated using a Vivaspin 15 (10,000 MWCO) centrifugal concentrator (Vivaproducts, Inc.).

Isothermal titration calorimetry

ITC measurements were performed on a low volume (300 µl sample cell) TA instruments NanoITC. Maltoheptaose (Sigma) solutions were prepared using the same dialysis buffer as the proteins, and thus protein and titrant were in the same buffer for all experiments. For each experiment 97 – 120 µM protein was placed in the sample cell. The reference cell was filled with deionized water. After the cell temperature was equilibrated to 25°C, an initial injection of 0.75 µl of maltoheptaose was performed followed by 27 subsequent injections of 1.75 µl of 1.67 – 7 mM maltoheptaose. The sample cell was stirred at 350 rpm and the resulting heat of reaction was measured. Data were analyzed using the NanoAnalyze software package (TA instruments) by fitting to an independent binding model and fixing N to the number of known binding sites in the protein

Growth competition experiments and Quantitative PCR (qPCR)

SusE/F strains used in this competition experiment were tagged with either pNBU2-tag 11 or 14 [42]. Three biological replicates of SusE/SusF strains were passaged each day for two weeks using a daily 1:100 back-dilution into MM containing 5 mg ml^{-1} glucose or 5 mg ml^{-1} maize amylopectin supplemented with 0.5 mg ml^{-1} maltose. Genomic DNA was harvested from cultures on days 0, 1, 2, 3, 4, 6, 8, 10, and 14 using DNeasy Blood & Tissue kit (Qiagen). Genomic DNA quantification was performed with a Mastercycler® ep realplex (Eppendorf), using KAPA SYBR® FAST qPCR Master Mix and 100 nM SusE and SusF primers, for 40 cycles of 95 °C for 3 s, 55 °C for 8 s, 72 °C for 20 s, followed by a melting step to determine amplicon purity. Samples were normalized to a DNA standard curve of genomic DNA from each respective strain. Relative abundance was calculated as the percent composition of a strain's DNA relative to the total DNA in the sample.

Western blotting

B. thetaiotaomicron strains were grown in MM containing 5 mg ml^{-1} maltose and were harvested at $O.D._{600}$=0.6. Cells were lysed in SDS sample buffer and 15 µg of total protein was loaded into an SDS-PAGE. SusF was detected in *B. thetaiotaomicron* whole cell lysates by western blot using custom rabbit polyclonal primary antibodies and horseradish peroxidase conjugated goat anti-Rabbit IgG secondary antibody (Sigma) [74].

Immunofluorescence

B. thetaiotaomicron strains were grown in 5 ml minimal *Bacteroides* medium supplemented with 5 mg ml^{-1} maltose to an $O.D._{600}$ of 0.6 and then harvested via centrifugation (7,000 x g for 3 min) and washed twice with phosphate- buffered saline (PBS). Cells were resuspended in 0.25 ml PBS, and 0.75 ml of 6% formalin in PBS was added. Cells were incubated with rocking at 20°C for 1.5 h and then washed twice with PBS. Cells were resuspended in 0.5 to 1 ml blocking solution (2% goat serum, 0.02% NaN_3 in PBS) and incubated for 16 h at 4°C. Cells were centrifuged and resuspended in 0.5 ml of a 1:100 dilution of custom rabbit anti-SusE or anti-SusF antibody sera in blocking solution and incubated by rocking for 2 h at 20°C. Cells were washed with PBS and then resuspended in 0.4 ml of a 1:500 dilution of Alexa Fluor 488 goat anti-rabbit IgG (Life Technologies) in blocking solution and incubated with rocking for 1 h at 20°C. Cells were washed three times with an excess of PBS and then resuspended in 20 µl of PBS plus 1 µl of ProLong Gold antifade (Life Technologies). Cells were spotted on 0.8% agarose pads and imaged using an Olympus IX70 inverted confocal microscope. Images were processed with Metamorph Software.

Thin layer chromatography

Cell-free culture supernatant was collected from stationary phase ΔSusG cultures grown in minimal media containing 5 mg ml^{-1} maltose. Supernatant was added to minimal media containing 5 mg ml^{-1} DP10 – 40 to make up 5% of the reaction volume. Reactions were incubated at 37°C for 1 hour and then flash frozen. These reactions were spotted onto TLC Silica gel 60 F254 20×20 cm glass plates (Millipore) and separated with the solvent acetonitrile:ethyl acetate:isopropanol:water

(40:10:25:525 until the solvent front was within 1 cm of the top of the plate. The sugars were then stained with 0.3% (w/v) N-(1- napthyl)ethylenediamine, 5% (v/v) sulfuric acid in methanol and heated until spots developed.

HPAEC-PAD Maltooligosaccharide Analysis

Samples were processed by the GlycoAnalytics Core at the University of California San Diego. Crude supernatant samples were passed over a PGC cartridge (poly-graphitized charcoal) HyperSep™ Hypercarb™ SPE Cartridges (Thermo Scientific), washed with 5 ml water and bound oligosaccharides were eluted with 30% Acetonitrile solution containing 0.1% TFA. This purification results in the loss of glucose and maltose. Eluted oligosaccharides were dried, resuspended in water, and injected on HPAEC-PAD. Oligosaccharide profiling was performed using BioLC CarboPac PA100 column (4X250mm) with PA100 (4x50 mm) guard column at a flow rate 1 ml min^{-1}. Pulsed amperometric detection with a gold electrode and standard quad waveform was used for carbohydrate analysis. The elution gradient was as follows: 0.0 – 20.0 min isocratic flow with 19 mM sodium hydroxide containing 7 mM sodium acetate, 20 – 70 min linear gradient of 0 mM – 400 mM sodium acetate. Maltrin-100 and Maltrin-200 were used as standards to compare the elution time for each oligosaccharide to verify the degree of polymerization. The area under each peak was calculated using Chromeleon™ 6.8 Chromatography Data System software and the DP values were assigned based on the retention time (min). Relative abundance is calculated as the percent composition of an oligosaccharide's peak area relative to the total area of all peaks in the sample.

Chapter III: The starch utilization system assembles around stationary starch-binding proteins

Notes

This chapter was reprinted and modified with permission from Tuson, H.H.,* Foley, M.F.*, Koropatkin, N.M., Biteen, J.S. The Starch Utilization System Assembles Around Stationary Starch-Binding Proteins. *Biophysical Journal*. **114**,1-9 (2018).

*These authors contributed equally

Abstract

Bacteroides thetaiotaomicron (*Bt)* is a prominent member of the human gut microbiota with an extensive capacity for glycan harvest. This bacterium expresses a five-protein complex in the outer membrane, called the Starch utilization system (Sus), which binds, degrades, and imports starch into the cell. Sus is a model system for the many glycan-targeting polysaccharide utilization loci found in *Bt* and other members of the Bacteroidetes phylum. Our previous work has shown that SusG, a lipidated amylase in the outer membrane, explores the entire cell surface, but diffuses more slowly as it interacts with starch. Here, we use a combination of single-molecule tracking, super-resolution imaging, reverse genetics, and proteomics to show that SusE and SusF, two proteins that bind starch, are immobile on the cell surface even when other members of

the system are knocked out and under multiple different growth conditions. This observation suggests a new paradigm for protein complex formation: binding proteins form immobile complexes that transiently associate with a mobile enzyme partner.

Introduction

The human gut microbiome comprises thousands of different bacterial strains that contribute to defense against pathogens, activation of the immune system, obesity, and malnutrition [109-113]. Many of these bacteria are required to break down complex plant-derived glycans that cannot be degraded by human digestive enzymes. Most glycan degradation depends on the collective action of the Gram-negative Bacteroidetes, a dominant bacterial phylum in the gut, which has evolved to degrade a vast array of carbohydrates [11]. This is accomplished through the regulated expression of outer membrane protein sets called starch-utilization system (Sus)-like systems, which are composed of several proteins that bind, degrade, and import different glycans. Sus-like systems are encoded within discrete polysaccharide utilization loci (PULs), and each Sus-like system outer membrane complex targets a distinct glycan substrate [15]. The repertoire of PULs within these bacteria drives their metabolic lifestyle and fitness, dictating both the colonization and persistence of these microbes within the host gut [40, 88]. A detailed understanding of the molecular mechanism of glycan transport is critical for the development of therapies that utilize diet to change the microbiota composition toward improved health.

Sus-like systems contain proteins homologous in structure and function to several proteins originally described in the prototypical PUL, the *Bacteroides*

thetaiotaomicron (*Bt)* Sus [24, 114]. The Sus locus (*susRABCDEFG)* encodes five outer membrane-localized proteins, SusCDEFG, which form the Sus complex at the cell surface (**Figure 3.1**). Starch polysaccharides are initially bound by SusD, SusE, and SusF, and are subsequently broken down into smaller oligosaccharides by the amylase SusG [26, 42, 56]. These oligosaccharides are transported into the periplasm through the SusC TonB-dependent transporter, and then degraded to glucose by the SusA and SusB glycosidases [31, 115].

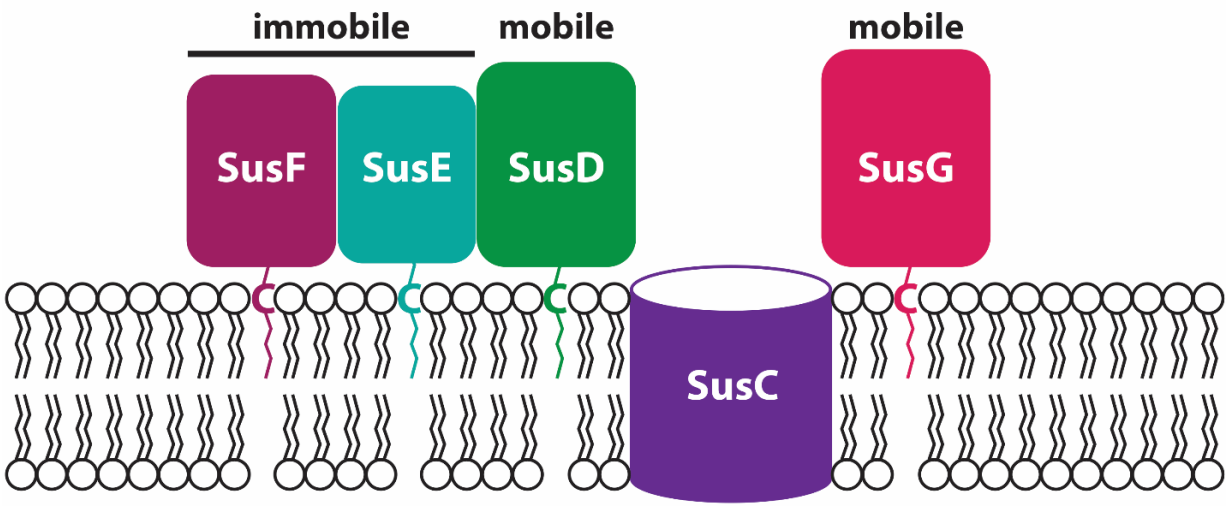

Figure 3.1. The outer membrane proteins of the Sus complex

SusC is a membrane-spanning β-barrel. SusDEFG are associated with the cell membrane through lipidation of a cysteine residue (C) near the N-terminus. The identity of the lipid groups is unknown. Single-molecule microscopy shows that SusD and SusG are mobile, while SusE and SusF are immobile.

The Sus is a model system for understanding how complex polysaccharides are broken down by Bacteroidetes, which rely heavily on Sus-like systems for nutrient acquisition [32, 39]. Nearly all sequenced gut Bacteroidetes genomes encode Sus-like systems that can metabolize a wide variety of substrates [15], yet little is known about how Sus proteins work together at the cell surface. The individual structures of the *Bt* Sus proteins SusCDEF have been elucidated, revealing that SusD has a single starch binding site, SusE and SusG each have two starch binding sites, and SusF has three starch binding sites [114]. These binding sites do not have redundant functionality, but rather discrete roles in glycan capture [69]. Importantly, SusDEFG are not embedded in the membrane, but rather tethered to the cell surface via lipidation at an N-terminal cysteine, followed by a flexible linker of 15 – 20 amino acids such that each protein can be thought of as "floating" above the membrane like a balloon on a string (**Figure 3.1**).

The SusDEFG proteins have discrete folds (**Figure 3.2**). SusD (62 kDa) is a globular *α*-helical protein while SusE (42 kDa) and SusF (52 kDa) are composed of three and four tandem Ig-like domains of ~100 amino acids, respectively; besides the N-terminal domain of each protein, all domains have a single starch-binding site. There is a proline residue in between nearly each domain junction, and we speculate that these prolines restrict the overall conformational flexibility of SusEF. SusG (78kDa) comprises a globular catalytic domain and a carbohydrate-binding module (CBM) that protrudes from the catalytic site. The CBM is linked to the catalytic domain by two short linkers (residues 212-217 and 334-338) that are unlikely to impart structural flexibility and thus SusG, like SusE and SusF, is expected to maintain an extended conformation. Additionally, the proteins lacking the signal peptides could be expressed in *Escherichia*

coli, were highly soluble, and bound starch. Therefore an interaction with a protein partner, or the *Bt* membrane, is not essential for folding or for starch binding. Still, despite the detailed information in **Figure 3.2** regarding the individual structures [42, 56, 74], we could not predict how the proteins behave in the membrane environment or how they interact with each other, or other proteins, on the cell surface.

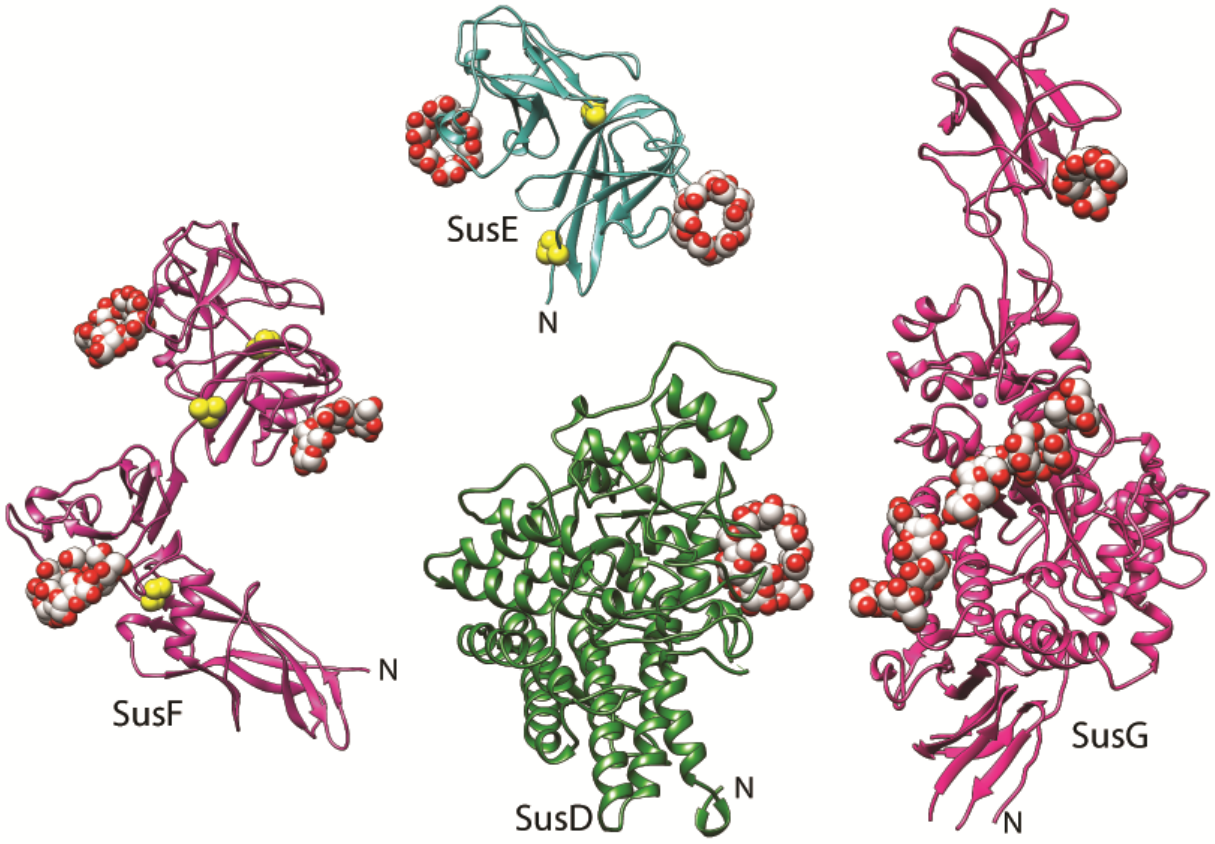

Figure 3.2. The crystal structures of SusD (PDB 3CK9), SusE (PDB 4FEM), SusF (PDB 4FE9), and SusG (PDB 3K8L)

Co-crystallized maltooligosaccharides are shown with grey and red spheres, prolines in SusE and SusF are shown with yellow spheres. The 15 – 20 residue N-terminal linkers that connect each protein to the lipidated cysteine for tethering to the membrane were not resolved in the crystal structures. The N-terminal domain of SusE (residues 38-167) was not resolved in the crystal structure.

We previously probed the movement of the glycoside hydrolase SusG in live anaerobic cells using single-molecule fluorescence imaging [79], and discovered that this protein moves around the entire cell surface. SusG displays a noticeable decrease in its effective diffusion coefficient in the presence of the large polymer starch, as well as when it transiently associates with the SusCDEF proteins, indicating dynamic, starch-mediated Sus complex formation. Little is known about the organization and structure of the *Bacteroides* membrane, and our work with SusG was the first to examine cell surface lipoprotein dynamics in live anaerobic cells. A key question is therefore whether other Sus surface components are similarly mobile, as answering this question would both inform a model of how the Sus proteins assemble in *Bt* and provide insight into the extent of protein mobility on the cell surface as bacteria interact with their environment.

Here we employ super-resolution imaging and single-molecule tracking to examine the starch-binding proteins SusE and SusF. In stark contrast to the highly mobile SusG, we find that SusE and SusF remain immobile at the cell surface despite changing environmental conditions and within strains with different genetic backgrounds. This is, to our knowledge, the first observation of lipoproteins that are significantly immobile (down to our 25 nm resolution). Similar high degrees of confinement have been observed in *E. coli* for some integral outer membrane proteins, with confinement diameters from published studies ranging from 0.02 to 0.60 μm [116, 117], though our observations here show that the SusE and SusF confinement diameter is *at most* 0.025 μm, i.e., less than or equal to the localization precision of our super-resolution microscope. This immobility does not appear to be driven by the N-terminal

sequence of the signal peptide or by lipidation as swapping in the N-terminal sequence of SusE does not decrease SusG diffusion. Therefore, we hypothesize that the differences in mobility are due to interactions of the SusEF proteins with an unknown partner, possibly another protein or a component of the cell membrane. Overall, we present here a new model for outer-membrane protein collaboration on the surface of Gram-negative bacteria: assembly of a system of mobile proteins around select stationary protein centers.

Materials and Methods

Bacterial strains, mutagenesis, and cell growth

Bacterial strains in this study are listed in **Table 3.1**. *Bacteroides thetaiotaomicron* (*Bt*) was grown as previously described [79]. Briefly, cells were cultured in medium containing 0.5% tryptone-yeast extract-glucose (TYG) and incubated at 37 °C under anaerobic conditions in a Coy chamber. Approximately 24 h prior to imaging, cells were diluted into minimal medium containing 0.5% (wt/vol) carbohydrate (glucose or maltose). Mutations were performed as previously described using a counter-selectable allelic exchange method [42]. All mutants were created in a thymidine kinase deletion (*Δtdk*) mutant. Oligonucleotides used in this study are listed in **Table A.2**.

Table 3.1. Bacterial strains used in Chapter III

Strain Name	Organism	Mutations	Notes
Bt Δ*tdk*	*Bt*	Δ*tdk*	Ref. (42)
Bt SusG-HT	*Bt*	SusG-HT	Ref. (80)
MF001	*Bt*	SusG-PamCherry, Δ*tdk*	
MF002	*Bt*	SusE-HT, Δ*tdk*	
MF003	*Bt*	SusE-PamCherry, Δ*tdk*	
MF004	*Bt*	SusE-PAmCherry Δ*susC*, Δ*tdk*	
MF005	*Bt*	SusE-PAmCherry Δ*susD*, Δ*tdk*	
MF006	*Bt*	SusE-PAmCherry Δcps	
MF007	*Bt*	SusF-HT, Δ*tdk*	
MF008	*Bt*	SusF-PAmCherry Δ*susC*, Δ*tdk*	
MF009	*Bt*	SusF-PAmCherry Δ*susD*, Δ*tdk*	
MF010	*Bt*	SusF-PAmCherry Δ*cps*, Δ*tdk*	

Imaging

Bt cells expressing fluorescent protein fusions were imaged on an Olympus IX71 inverted fluorescence microscope using a 1.40 numerical aperture 100x widefield oil-immersion objective. HaloTag labeling with TMR was performed as described previously [79], and TMR-labeled fusions were excited with a 561-nm laser (Coherent-Sapphire 561-50). Fusions to PAmCherry were activated with a 406-nm laser (Coherent Cube 405-100) and excited with the 561-nm laser. Fluorescence emission was detected using a 512 × 512 pixel Photometrics Evolve electron-multiplying charge-coupled device camera. Images were collected continuously at a rate of 25 frames per second (fps) or at 1 fps for time-lapse imaging.

Data analysis

Single-molecule fitting and track construction were performed as previously described [118]. The mean squared displacement (MSD) was calculated for each track, then trimmed so that only the first half of the MSD was retained to reduce the noise produced by averaging fewer steps at longer time lags (τ). MSD vs. τ curves of length 10 or longer were used for further analysis (i.e., the original track had to be at least 20 steps long to be considered). The slope of each MSD curve was calculated by fitting a line to the first three points.

Western blotting, co-IP, and membrane fractionation

Bt membranes were collected by spinning from mid-log cells grown anaerobically in minimal medium containing 0.5% (w/v) maltose. Crosslinking was carried out anaerobically for 1 h on whole cells before fractionation. Cells were lysed by sonication and spun briefly to remove large debris. Membranes were collected by spinning for 30 min at 50,000 x g, washed with phosphate-buffered saline (PBS), then resuspended and incubated in PBS containing 1.5% (w/v) dodecyl maltoside. Solubilization was performed for 1 h, and remaining insoluble protein was spun for 45 min at 100,000 x g. Co-IP was performed on solubilized membranes using the Pierce Crosslink Magnetic IP/Co-IP Kit according to manufacturer instructions. Sus outer membrane lipoproteins were detected by Western blot using rabbit polyclonal primary antibodies and horseradish peroxidase conjugated goat anti-Rabbit IgG secondary antibody (Sigma). Antibodies to SusD, SusE and SusF were raised against the recombinantly expressed proteins [69, 119]. The SusC antibodies were raised against the N-terminal plug domain of the protein (residues 118-242), which is conserved among many other SusC-like

proteins in *Bt*. Therefore, the SusC antibodies cross-react with other SusC-like proteins present in our membrane preparations, as seen in **Figure 3.12B**.

Total membrane proteomic sample preparation

1 L of *Bt* was grown to mid-log phase (OD_{600} 0.65 – 0.75) at 37 °C in minimal *Bacteroides* medium with 5 mg/mL maltose as the carbon source. Cells were centrifuged (10,000 x g for 15 min) then frozen in liquid N_2. Cells were thawed in 20 mL of 20 mM KH_2PO_4 (pH 7.3), sonicated on ice, and intact cells were removed by centrifugation at 15,000 x g for 17 min. The supernatant was ultracentrifuged at 200,000 x g for 2 h at 4 °C to pellet total membranes. The membrane pellet was resuspended in 20 mL of 20 mM KH_2PO_4 (pH 7.3), followed by a second round of ultracentrifugation. All steps from growth to membrane purification were repeated, and both preparations were submitted for mass spectrometry (runs 1 and 2; **Table A.5**). For the first run, the final membrane pellet was resuspended in 7.5 mL of 20 mM KH_2PO_4 (pH 7.3) plus 0.1% Tween-20 and gently sonicated to completely resuspend the membranes. This sample was concentrated via a 5 kDa MWCO spin filter prior to proteomic analysis. For the second preparation only, half a tablet of cOmplete EDTA-free protease inhibitor (Roche) was included during cell lysis and the final membrane pellet was resuspended in 3 mL of 20 mM KH_2PO_4 (pH 7.3) plus 0.1% Tween-20; this sample did not require concentration.

Quantitative Mass Spectroscopy

Bt total membrane samples and co-IP samples were submitted to MS Bioworks LLC (Ann Arbor, MI) for Quant-works Label-free Unfractionated proteomic analysis, as described in ref. [14]. The protein abundances within each sample were determined by calculating for each protein the normalized spectral abundance factor, $NSAF = (SpC/MW)/\sum_i(SpC_i/MW_i) \cdot N$, where *SpC* is spectral counts, *MW* is the protein molecular weight, and *N* is the total number of proteins.

Bt growth experiments

All *Bt* culturing was performed in a Coy anaerobic chamber (gas mix: 85% N_2, 10% H_2, 5% CO_2) at 37°C. Each strain was grown for 16 h from a freezer stock in tryptone-yeast extract-glucose (TYG) medium and then back-diluted 1:100 into *Bacteroides* minimal medium supplemented with 5 mg/mL glucose. After 24 h, cultures were back-diluted 1:100 into *Bacteroides* minimal medium supplemented with 5 mg/mL of carbohydrate. Growth experiments were performed in triplicate. Plates were loaded into a Biostack automated plate-handling device coupled to a Powerwave HT absorbance reader (both from Biotek Instruments, Winooski, VT). Absorbance at 600 nm (OD_{600}) was measured for each well at 20 min intervals. Data were processed using Gen5 software (BioTek), Microsoft Excel, and Prism.

Immunofluorescence

Bt strains were grown in *Bacteroides* medium with 5 mg/mL maltose to an OD_{600} of 0.6 and washed twice with PBS. Cells were resuspended in 0.25 mL PBS, to which 0.75 mL of 6% (v/v) formalin in PBS was added, and then incubated with rocking at

20°C for 1.5 h. Cells were washed twice with PBS, then resuspended in 0.5 – 1 mL blocking solution (2% (v/v) goat serum, 0.02% (w/v) NaN_3 in PBS) and incubated for 16 h at 4 °C. Cells were centrifuged and resuspended in 0.5 mL of a 1/100 dilution of custom rabbit antibody sera in blocking solution and incubated by rocking for 2 h at 20°C. Cells were washed with PBS and then resuspended in 0.4 mL of a 1/500 dilution of Alexa Fluor 488 goat anti-rabbit IgG (Life Technologies) in blocking solution and incubated with rocking for 1 h at 20 °C. Cells were washed three times with an excess of PBS and then resuspended in 20 µL of PBS plus 1 µL of ProLong Gold antifade (Life Technologies). Cells were spotted on 0.8% (w/v) agarose pads and imaged using an Olympus IX70 inverted confocal microscope. Images were processed with Metamorph Software.

Results and Discussion

SusG dynamics are unchanged under aerobic conditions

Although we have previously shown that fusions of SusG to the enzymatic HaloTag can be labeled by the dye TMR for single-molecule imaging in living, anaerobic *Bt [79]*, the fluorescent protein PAmCherry is advantageous because its photoactivation allows control of the number of molecules that are fluorescent at one time. *Bt* is aero-tolerant, and colonies are able to survive oxygen exposure for several days [120]. We explored the possibility of imaging dormant cells under aerobic conditions to allow the use of PAmCherry, whose chromophore requires oxygen to mature. We found that SusG-HaloTag is mobile under aerobic conditions (**Figure 3.3 A,B**). The single-molecule trajectories of the proteins can be quantitatively analyzed in plots of the mean

squared displacement (MSD) vs. time lag (**Figure 3.3C,D**). Here, each trajectory is a curve with slope equal to 4 times the effective molecular diffusion coefficient, *D*. SusG-HaloTag protein dynamics are unchanged in the presence or absence of oxygen (**Table 3.3**). As before, we constructed SusG-PAmCherry by replacing the starch-binding CBM58 domain with PAmCherry [79]. The SusG-PAmCherry fusion supports growth on starch (**Figure 3.4**), and similar effective diffusion coefficients and distributions were measured for SusG-PAmCherry in aerobic conditions as for the SusG-HaloTag fusion under anaerobic conditions (**Table 3.2**; ref. [79]). The dynamics of SusG are therefore independent of both the identity of the fluorescent tag and the presence or absence of oxygen in the environment. The similar patterns of diffusion in aerobic and anaerobic conditions imply that the SusG dynamics are passive and likely a result of membrane fluidity, rather than dependent upon the cell's redox status or on the proton motive force, as oxygen prevents maltooligosaccharide uptake by *Bt*. Because SusE and SusF are lipoproteins like SusG, and their glycan-binding activity is not oxygen-sensitive, we examined the dynamics of these proteins using PAmCherry fusions and in aerobic conditions.

Table 3.2. Dynamics of protein constructs

Distributions of effective diffusion coefficients, *D*, for each construct measured with single-molecule tracking. Values for *D* are in $\mu m^2/s$.

Protein Construct	Aerobic/ Anaerobic	Percent tracks with $D \leq 0.001$	Percent tracks with $0.001 < D \leq 0.02$	Percent tracks with $0.02 < D \leq 0.05$	Percent tracks with $0.05 < D \leq 0.1$	Percent tracks with $D > 0.1$	Total tracks	Tracks shown in Fig.
SusG-HT	anaerobic	25.7	52.1	17.4	3.3	1.5	1687	S2d
SusG-HT	aerobic	29.3	52.7	12.1	4.0	2.0	5788	S2c
SusG-PAmCherry	aerobic	29.3	44.1	14.4	8.6	3.5	256	2f
SusE-PAmCherry	aerobic	43.2	50.2	5.5	1.1	none	273	2d
SusE-HT	anaerobic	53.4	43.2	2.7	none	2.1	146	S6f
SusE-HT	aerobic	43.3	48.7	6.3	1.4	0.3	348	S6e
SusF-PAmCherry	aerobic	46.5	48.8	3.9	0.6	0.3	338	2e
SusD-PAmCherry (3-Ala linker)	aerobic	21.8	55.5	16.4	4.6	1.8	110	S7c
SusD-PAmCherry (20-Ala linker)	aerobic	23.6	54.7	15.1	3.8	2.8	106	S7d
SusE-Nterm-SusG-PAmCherry	aerobic	23.8	42.9	15.5	11.9	5.6	84	4d

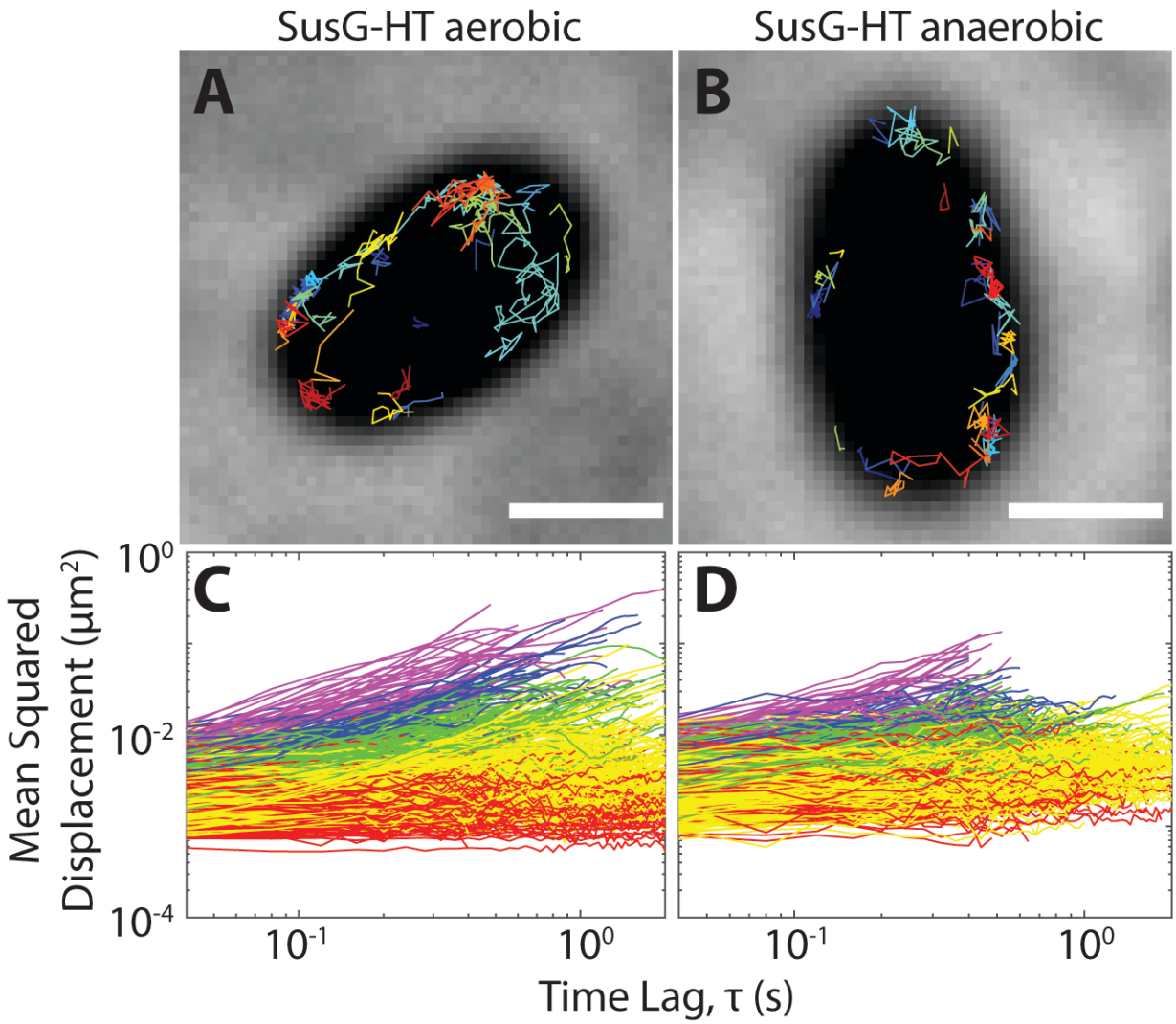

Figure 3.3. SusG-HT is mobile on the cell surface whether imaged under aerobic or anaerobic conditions

(A-B): each image shows a cell with 35 SusG-HT single-molecule tracks plotted in random colors. Scale bars = 1 μm. (C-D): the mean squared displacement of all tracks lasting longer than 20 frames is plotted for each protein fusion. Red: effective diffusion coefficient (D) ≤ 0.001 $\mu m^2/s$; yellow: 0.001 $\mu m^2/s < D \leq 0.02$ $\mu m^2/s$; green: 0.02 $\mu m^2/s < D \leq 0.05$ $\mu m^2/s$; blue: 0.05 $\mu m^2/s < D \leq 0.1$ $\mu m^2/s$; purple: $D > 0.1$ $\mu m^2/s$.

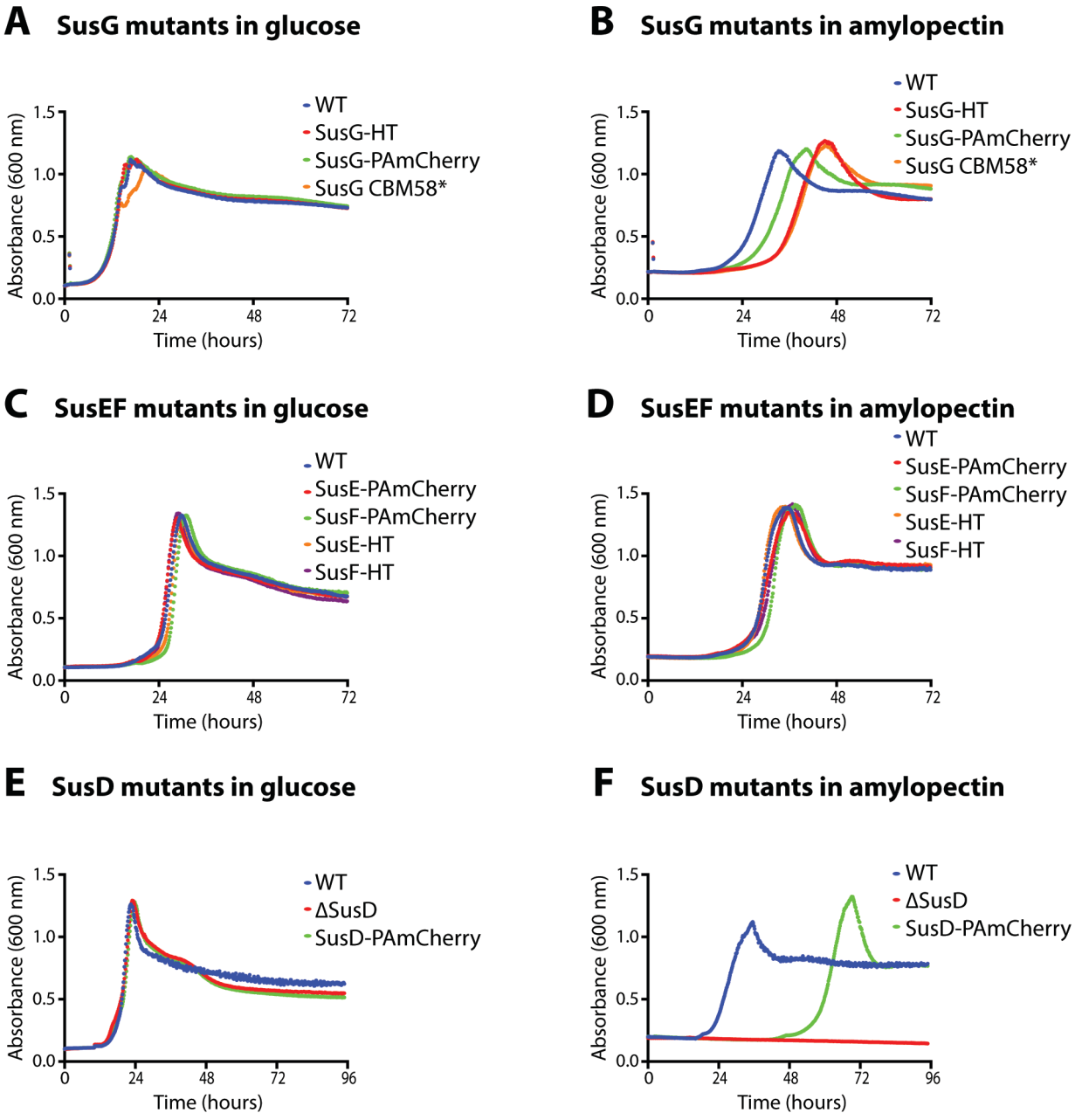

Figure 3.4. The fluorescently labeled SusG strains support growth on starch

Growth was measured in minimal medium containing 5 mg/mL glucose (*left*) or maize amylopectin (*right*) as the sole carbon source. (A, B): SusG-HT and SusG-PAmCherry were made by replacing the starch-binding CBM58 domain with HT or PAmCherry; CBM58 is not required for growth on starch as evidenced by the normal growth of SusG CBM58*, which contains a starch-bindingdeficient version of CBM58. (C, D): SusE-HT, SusE-PAmCherry, SusF-HT, and SusF-PAmCherry are C-terminal protein fusions with 3-alanine linkers. (E, F): SusD-PAmCherry is a C-terminal protein fusion with a 20-alanine linker. The ΔSusD strain was used as a negative control for growth on starch.

SusE and SusF are immobile in the membrane despite perturbations to the cellular environment

PAmCherry was fused to the C-terminus of SusE or SusF via a 3-alanine linker. Cells expressing these fusions displayed wild-type growth kinetics on starch (**Figure 3.4C,D**), indicating that the fusions are functional. Immunofluorescence on fixed, non-permeabilized *Bt* cells using antibodies against SusE and SusF demonstrated that PAmCherry-labeled SusE and SusF localize to the outer surface of the cell as expected (**Figure 3.5**). To our surprise and in contrast to our SusG fusions, the single-molecule trajectories of the SusE and SusF fusions are compact and confined to puncta, indicating that these constructs are immobile in the cell membrane over time spans ranging from a few seconds to tens of seconds (**Figure 3.6A,B, 3.7, Movie B.1, B.2**). Of note, we observe mobile SusG and immobile SusE and SusF in identically treated cells, supporting that these dynamic observations are not due to unintended differences in the experimental conditions. The MSD curves show clear differences for the various Sus lipoproteins (**Figure 3.5D,F**, **Table 3.2**). While 12.1% of SusG-PAmCherry molecules diffuse at a rate greater than 0.05 $\mu m^2/s$ (blue and purple curves in **Figure 3.5F**), only 1.1% of SusE-PAmCherry and 0.9% of SusF-PAmCherry move this quickly (blue and purple curves in (**Figure 3.5E,F**). Similarly, while 93.4% of SusE-PAmCherry and 95.3% of SusF-PAmCherry molecules have effective diffusion coefficients < 0.02 $\mu m^2/s$ (red and yellow curves in **Figure 3.5D,E**), only 73.4% of SusG-PAmCherry molecules move this slowly (red and yellow curves in **Figure 3.5F**). To further confirm that Sus protein dynamics are unaffected by the presence of oxygen, we fused SusE to the HaloTag (HT) protein and imaged TMR-labeled SusE-HT under aerobic and anaerobic

conditions. Like SusE-PAmCherry imaged aerobically, SusE-HT remained confined to puncta under both conditions (**Figure 3.8**) and had a similar mobility profile (**Table 3.2**). The similarities within these different labeling and oxygen conditions further demonstrate that anaerobic conditions are not required to obtain physiologically relevant protein dynamics.

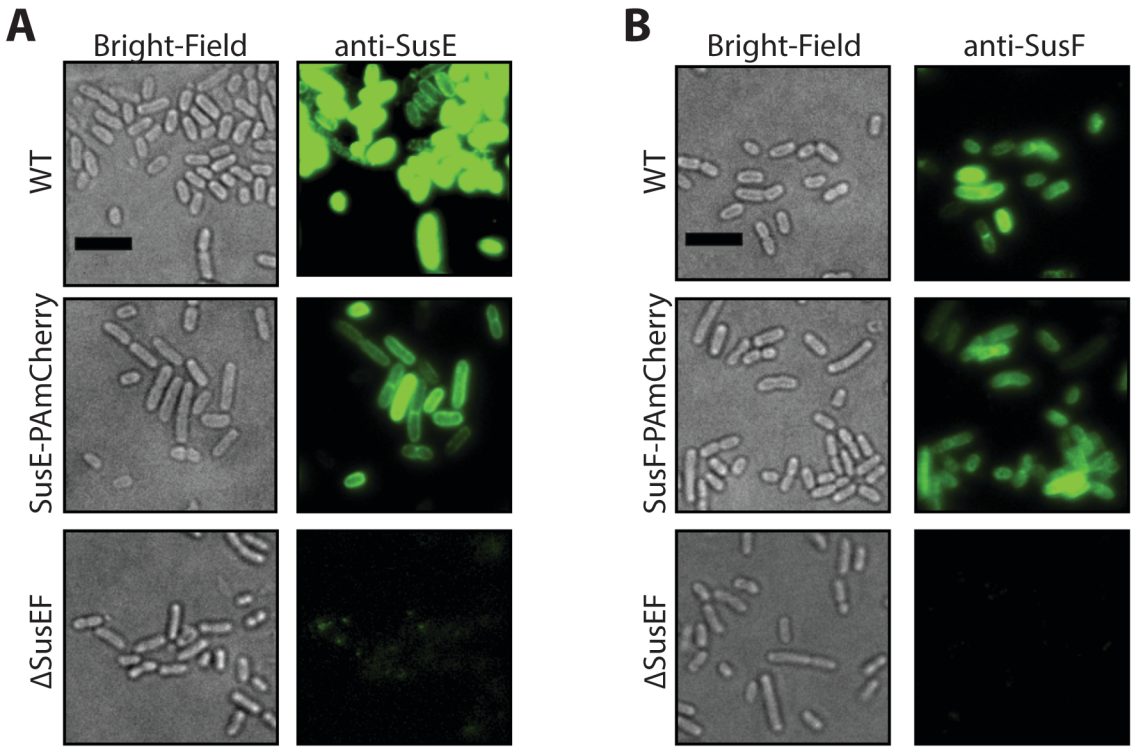

Figure 3.5. SusE-PAmCherry and SusF-PAmCherry visualized by immunofluorescence

Formalin-fixed, non-permeabilized *Bt* strains were grown in minimal media supplemented with maltose and labeled with custom rabbit polyclonal antibodies to SusE and SusF and then stained with a secondary antibody conjugated to Alexa Fluor 488 goat anti-rabbit IgG. The side-by-side panels display bright-field and fluorescence images for each strain labeled with (A) anti-SusE serum and (B) anti-SusF serum. Scale bars = 5 μm.

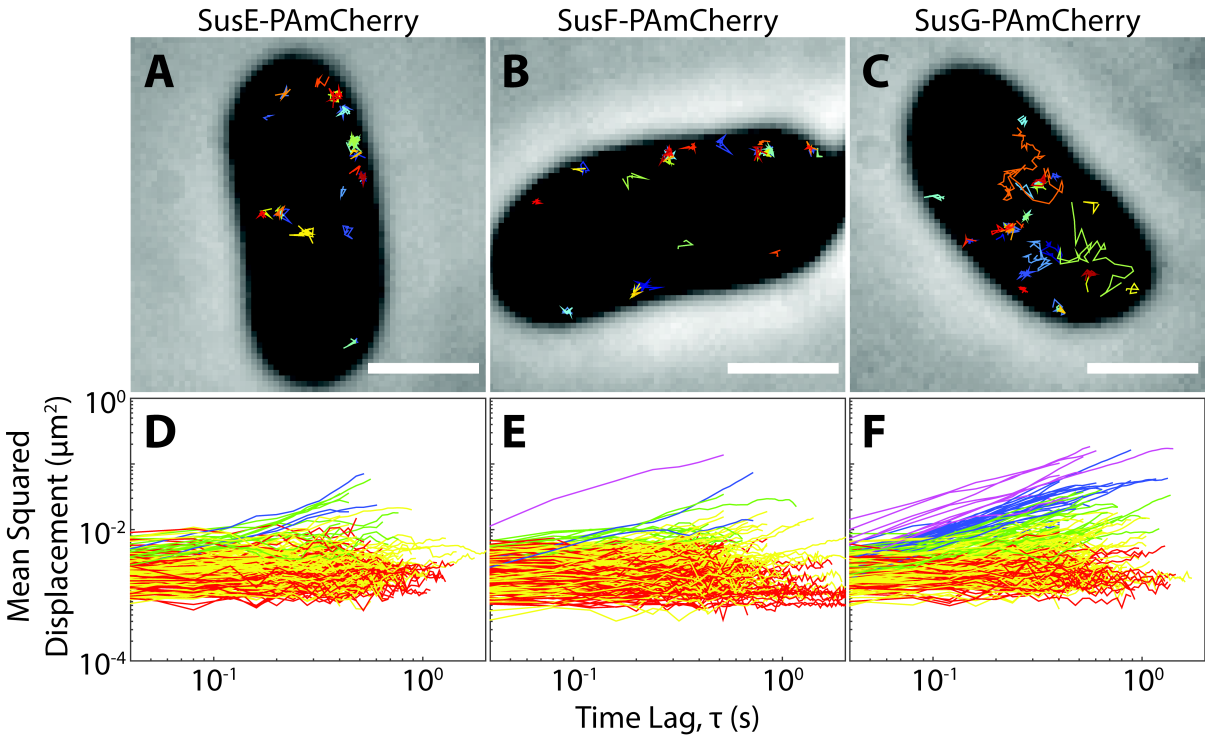

Figure 3.6. SusE-PAmCherry and SusF- PAmCherry are highly confined, whereas SusG- PAmCherry explores the cell.

(A–C) Each image shows a cell with 35 single-molecule tracks plotted in random colors. Scale bars, 1 mm. (D–F) MSD of all tracks lasting >20 frames is plotted for each protein fusion (red, effective diffusion coefficient (D) % 0.001 mm^2/s; yellow, 0.001 mm^2/s < D % 0.02 mm2/s; green, 0.02 mm^2/s < D % 0.05 mm^2/s; blue, 0.05 mm^2/s < D % 0.1 mm^2/s; purple, D > 0.1 mm^2/s).

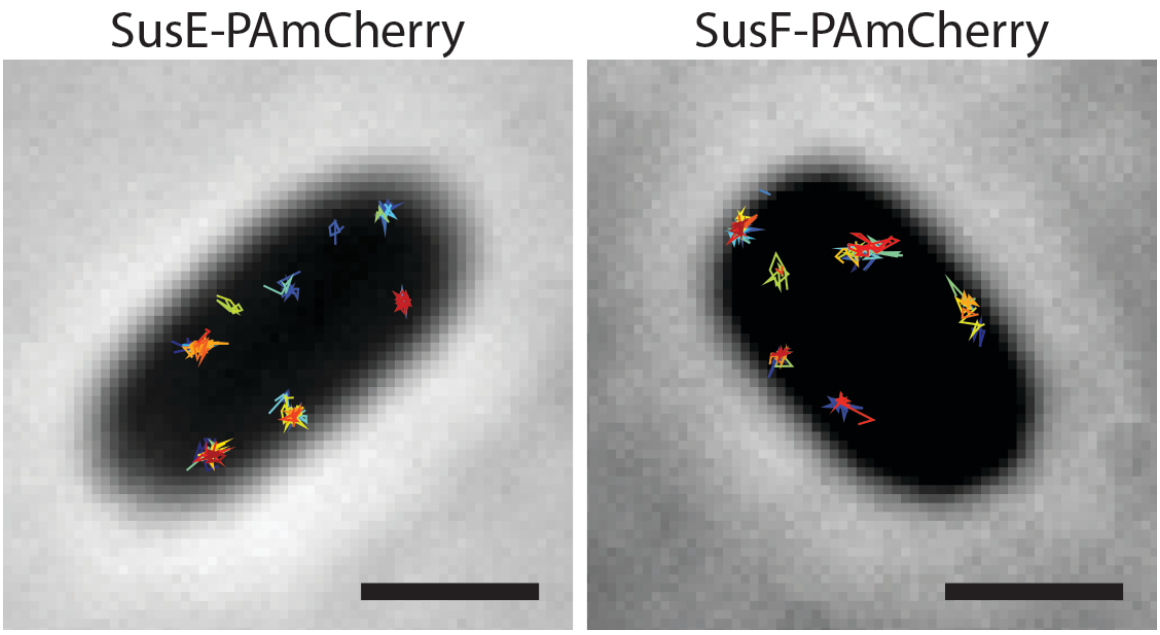

Figure 3.7. SusE-PAmCherry and SusF-PAmCherry remain highly confined when imaged for longer periods of time via time-lapse imaging

These figures show tracks of molecules that remain in place for 3 – 47 s, with one 40 ms frame acquired every 1 second. Each image shows a cell with 35 single-molecule tracks plotted in random colors. See also the corresponding **Movies A.1** and **A.2** of SusE-PAmCherry and SusF-PAmCherry, respectively. Scale bars = 1 μm.

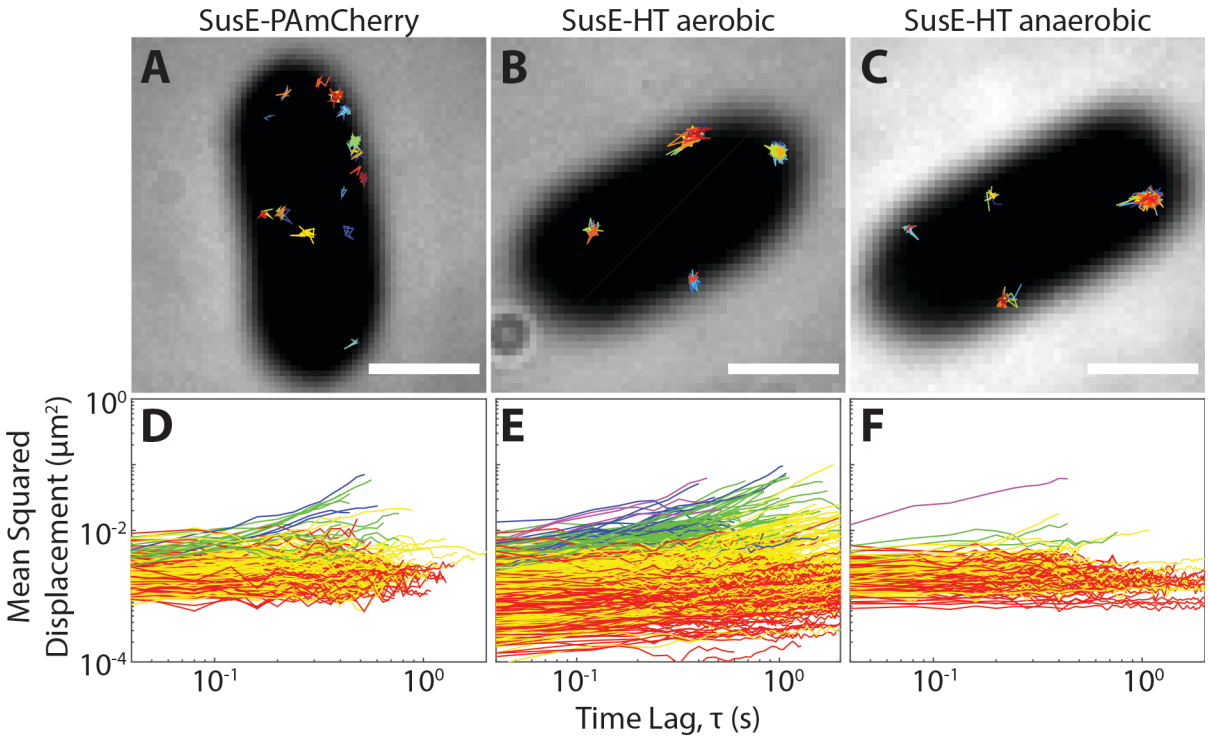

Figure 3.8. SusE-PAmCherry is highly confined when labeled with PAmCherry or HaloTag (HT) and whether imaged under aerobic or anaerobic conditions

(A-C): each image shows a cell with 35 single-molecule tracks plotted in random colors. Scale bars = 1 μm. (D-F): the mean squared displacement of all tracks lasting longer than 20 frames is plotted for each protein fusion. Red: effective diffusion coefficient (D) ≤ 0.001 μm^2/s; yellow: 0.001 μm^2/s < D ≤ 0.02 μm^2/s; green: 0.02 μm^2/s < D ≤ 0.05 μm^2/s; blue: 0.05 μm^2/s < D ≤ 0.1 μm^2/s; purple: D > 0.1 μm^2/s.

We hypothesize that the enzymatic SusG diffuses in the outer membrane while SusE and SusF remain stationary for optimal carbohydrate binding by the Sus system. To complete the picture, we therefore examined the mobility of a fusion of PAmCherry to SusD, the fourth carbohydrate-binding Sus outer membrane protein. SusD-PAmCherry exhibited high mobility like SusG-PAmCherry (**Figure 3.9, Table 3.2**). However, it should be noted that the SusD-PAmCherry strain exhibited delayed growth on starch, which suggests that this tag reduced the ability to SusD to effectively contribute to starch uptake. Therefore, it is difficult to discern if the movement of SusD is due to the tag interfering with the interaction of SusD with its partner SusC, or if SusD is normally mobile. Changing the length of the linker between the C-terminus of SusD and PAmCherry from a 3-Ala linker to a 20-Ala linker did not resolve the growth defect (**Figure 3.4E,F**).

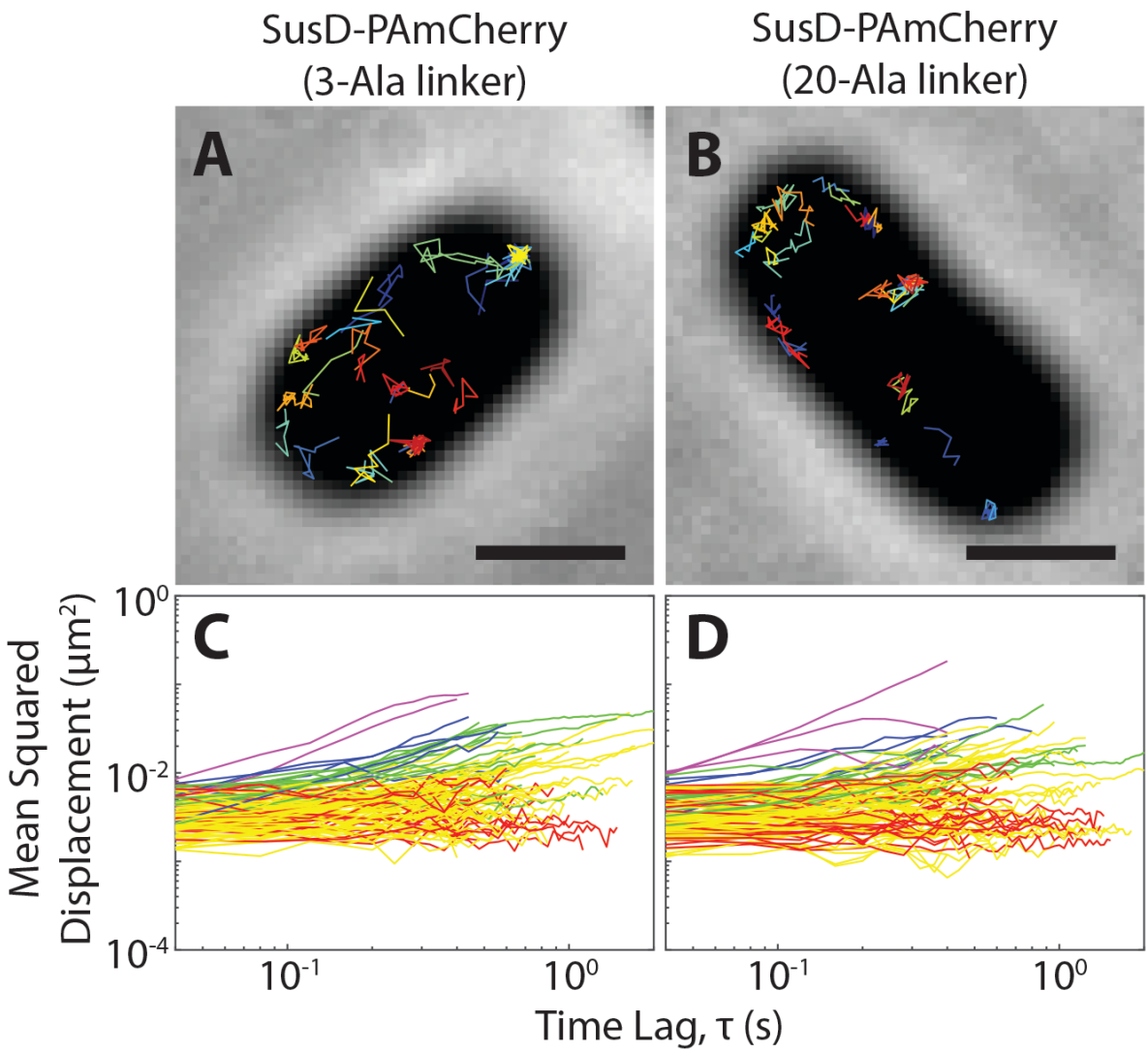

Figure 3.9. SusD-PAmCherry is mobile when the fluorescent label is attached via two different C-terminal linker lengths

(A, B): each image shows a cell with 35 single-molecule tracks plotted in random colors. Scale bars = 1 μm. (C, D): the mean squared displacement of all tracks lasting longer than 20 frames is plotted for each protein fusion. Red: effective diffusion coefficient (D) ≤ 0.001 μm^2/s; yellow: 0.001 μm^2/s < D ≤ 0.02 μm^2/s; green: 0.02 μm^2/s < D ≤ 0.05 μm^2/s; blue: 0.05 μm^2/s < D ≤ 0.1 μm^2/s; purple: D > 0.1 μm^2/s.

Because SusE and SusF remain stationary in the outer cell membrane despite changes in oxygen and labeling conditions, we further measured the effect of the presence of other Sus proteins and capsular polysaccharide on the positioning of SusEF. Previous studies with formaldehyde crosslinking have demonstrated that SusC, SusD, and SusE interact [27], so we hypothesized that SusCD might be responsible for confining SusE and SusF in the membrane. However, both SusE and SusF remained immobile when SusC or SusD were deleted from the chromosome (**Figure 3.10A-D**). Furthermore, SusF remained immobile in a strain in which SusE was deleted (**Figure 3.10G**). Thus, the stationary character of SusE does not depend on SusCD, and the stationary character of SusF does not depend on SusCDE. *Bt* has a thick surface capsular polysaccharide layer [15]. Although the precise monosaccharide and linkage composition of these capsules is unknown, we hypothesized that the glycan-binding proteins SusE and SusF may interact nonspecifically with the capsule or are otherwise influenced by the capsule organization on the cell surface. We examined SusEF dynamics in an acapsular strain of *Bt*, and found that both SusE-PAmCherry and SusF-PAmCherry remain immobile in the absence of the capsule (Δ*cps*; **Figure 3.10E,F**). This is consistent with our previous demonstration that SusG protein dynamics are unchanged by the presence or absence of capsule [79]. Thus, SusEF interactions with the capsule layer are not responsible for immobility.

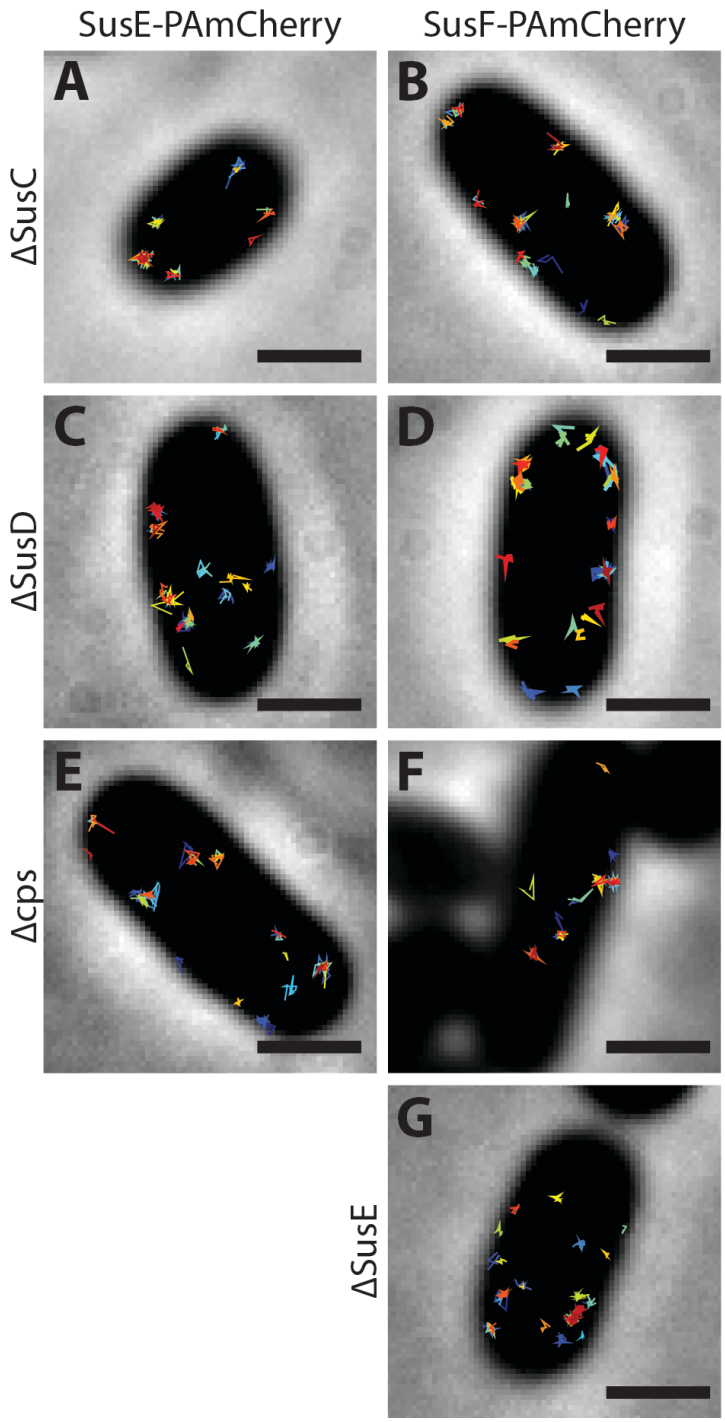

Figure 3.10. SusE-PAmCherry and SusF-PAmCherry remain highly confined even when other members of the Sus complex or the capsule (cps) machinery are knocked out

Each image shows a single cell with 35 single-molecule tracks plotted in random colors. Scale bars = 1 μm.

The SusE N-terminus does not confer immobility

Following cleavage of the signal peptide, mature SusDEFG are lipidated at an N-terminal cysteine and tethered to the outer leaflet of the outer membrane [27, 28]. Previous structural studies of these lipoproteins demonstrate that they also contain a flexible disordered linker that further separates the N-terminal domain from the surface of the cell [114]. To determine if the N-terminal sequence of SusE confers immobility, we replaced the N-terminal region (M1-W44) of SusG—which contains the signal peptide through the flexible linker—with the analogous N-terminal region (M1-N28) of SusE. This "SusE-Nterm-SusG-PAmCherry" hybrid protein remained mobile, with dynamics indistinguishable from those of SusG-PAmCherry (**Figure 3.11, Table 3.2**). 73.4% of SusG-PAmCherry molecules and 66.7% of SusE-Nterm-SusG-PAmCherry molecules diffuse slower than 0.02 $\mu m^2/s$; 14.4% of SusG-PAmCherry molecules and 15.5% of SusE-Nterm-SusG-PAmCherry molecules diffuse between 0.02 and 0.05 $\mu m^2/s$; and 12.1% of SusG-PAmCherry molecules and 17.5% of SusE-Nterm-SusG-PAmCherry molecules diffuse faster than 0.05 $\mu m^2/s$. These results suggest that the immobility of SusE is not conferred by its N-terminus, but is likely due to an unknown interaction with some membrane component or another protein. However, treatment of *Bt* cells with ethylenediaminetetraacetic acid (EDTA) or lysozyme to disrupt lipopolysaccharide or peptidoglycan, respectively, failed to increase SusEF mobility (data not shown).

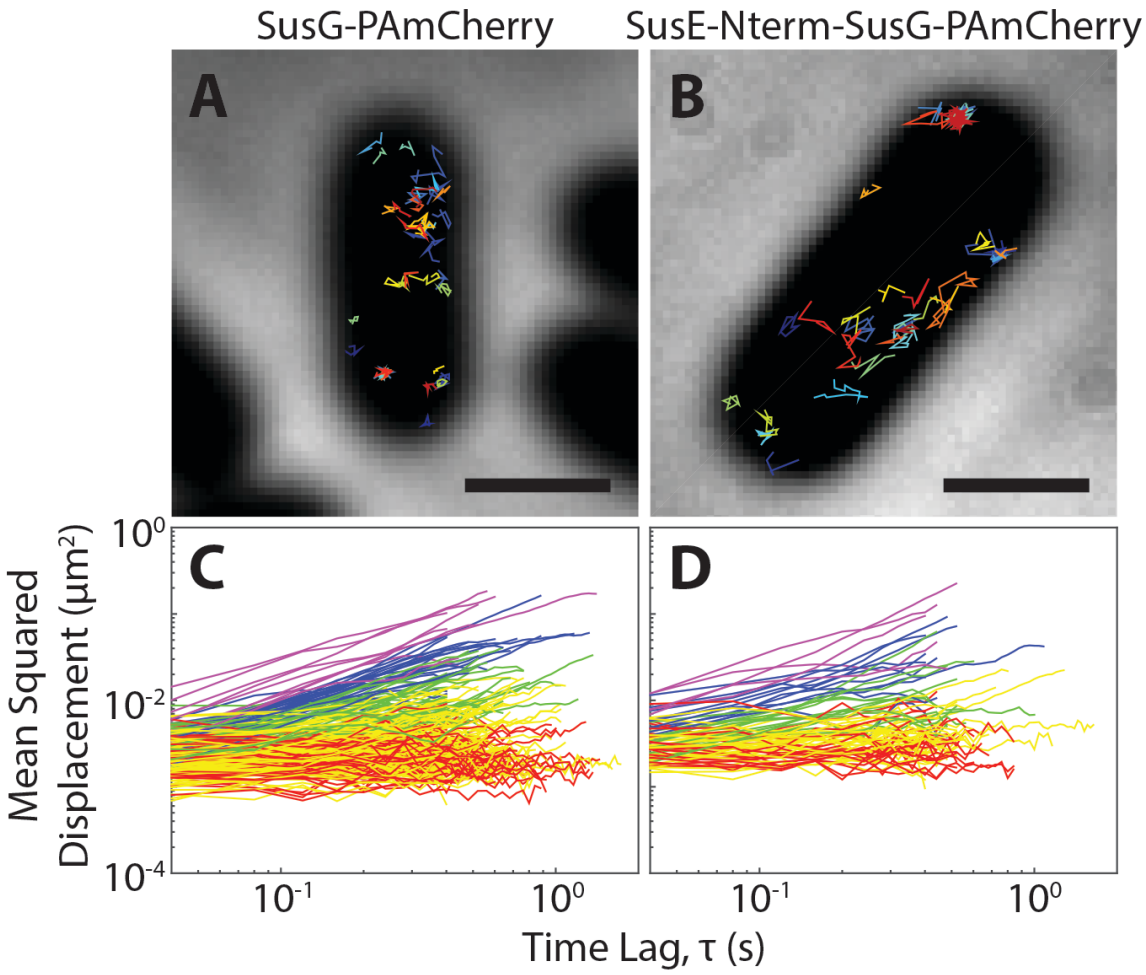

Figure 3.11. SusG-PAmCherry remains mobile when the 44 N-terminal amino acids are replaced with the 28 N-terminal amino acids from SusE changing the lipidation signal

(A, B) each image shows a cell with 35 single-molecule tracks plotted in random colors. Scale bars = 1 μm. (C, D) the mean squared displacement of all tracks lasting longer than 20 frames is plotted for each protein fusion. Red: effective diffusion coefficient (D) ≤ 0.001 μm^2/s; yellow: 0.001 μm^2/s < D ≤ 0.02 μm^2/s; green: 0.02 μm^2/s < D ≤ 0.05 μm^2/s; blue: 0.05 μm^2/s < D ≤ 0.1 μm^2/s; purple: D > 0.1 μm^2/s.

Overall, though our experiments do not identify the mechanism that immobilizes SusEF, they demonstrate that the immobility of SusEF in the *Bt* outer membrane is an important property that remains robust to perturbations: SusEF remain stationary in the membrane independent of fluorescent tag identity or oxygen concentration and despite numerous perturbations, including knockouts of other Sus system proteins. This strong propensity of several proteins to remain stationary while their putative binding partners are mobile in the bacterial cell membrane presents a new model for the cooperative action of an outer membrane protein system.

SusCDE can be captured in an outer membrane complex

Previous work has captured a SusCDE interaction via formaldehyde crosslinking of cells followed by native polyacrylamide gel electrophoresis (PAGE) [27]. To further examine this interaction and any others that may take place within this complex, we performed co-immunoprecipitation (co-IP) of SusD using custom anti-SusD with solubilized *Bt* membranes. This solubilization efficiently releases SusD, SusE, and SusG; however, SusF remains in the insoluble fraction (**Figure 3.12A**), even upon prolonged incubation with detergent (**Figure 3.13**). SusF is an otherwise soluble protein when expressed recombinantly without its lipidation site, so its enrichment in the insoluble membrane fraction suggests that it interacts strongly with the membrane either directly or via binding to an unknown insoluble molecule. While it is tempting to speculate that the immobility of both SusE and SusF is due to their interactions in the membrane, SusE is mostly solubilized while SusF is not. Thus, it is unclear whether the

enrichment of SusF in the insoluble fraction is related to its stationary nature, or whether these two observations are independent characteristics of this protein.

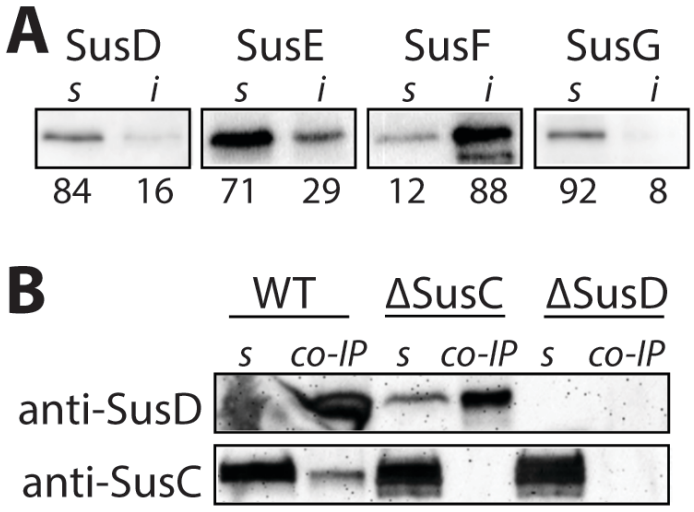

Figure 3.12. The Sus outer membrane proteins vary in solubility and abundance

(A) The soluble (*s*) and insoluble (*i*) fractions of Sus outer membrane lipoproteins after detergent extraction from fractionated *Bt* membranes. The relative density, as a percent of the total signal, is reported below each lane. (B) Co-IP of SusD from solubilized membranes of the three strains. Western blot on the Co-IP samples was performed with custom SusD or SusC antibodies. Of note, anti-SusC cross-reacts to label other SusC-like proteins in *Bt* (Methods). The soluble (*s*) lane is the solubilized membranes used for the co-IP, and the *co-IP* lane indicates the sample that was eluted after immunoprecipitation.

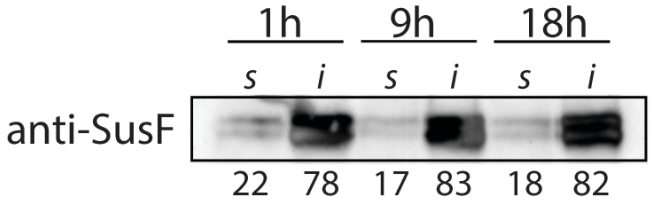

Figure 3.13. SusF remains insoluble during prolonged incubation with dodecyl maltoside

SusF is solubilized from the membrane as described in Methods, but incubated for 1, 9 or 19 h prior to centrifugation. The membrane-solubilized supernatant (*s*) and insoluble material (*i*) were run on a Western blot and SusF was detected with custom anti-SusF rabbit antibodies.

Based on previous reports, we hypothesized that some of the Sus proteins would interact through transient or weak protein-protein interactions [27]. To detect Sus complex formation, we performed quantitative proteomics on the total membrane fraction from *Bt* prior to solubilization, as well as on SusD and SusE co-IP elutions from cells with or without formaldehyde crosslinking (**Table 3.3, Table A.6, A.7, A.8, A.9, A.10**). For these experiments, we report the normalized spectral abundance factor (*NSAF*), a measure of how much of the total spectrum for a sample can be assigned to a particular protein, normalized for the molecular weight of that protein [121]. Under native (non-crosslinking) conditions, co-IP of SusD brings along SusC, as demonstrated via Western blot (**Figure 3.12B**), and the relative abundance of SusD:SusC is approximately 1:0.6 (**Table 3.3, A.6, A.7**). However, very little SusE is captured as a co-eluent in the native SusD pull downs; the relative abundance of SusD:E is at most 1:0.1. This low ratio of SusD:SusE, and the fact that it was only detected in one native experiment, suggests that SusE interacts only weakly with SusD or SusC in this experiment. When *Bt* cells were treated with formaldehyde to covalently link weakly or transiently interacting proteins, more SusE was immunoprecipitated along with SusD. For reasons that remain unclear, a lower amount of SusC was captured with SusD when formaldehyde was added. It is possible is that SusE and SusC compete for interactions with SusD such that capturing more SusE limits the ability of SusC to interact with SusD. Reciprocally, co-IP of SusE under native conditions did not capture additional Sus proteins, but formaldehyde treatment resulted in co-elution of SusD and SusC (**Table 3.3, A.10**). This need for formaldehyde suggests that the interaction of SusE with the SusCD complex is weak and/or transient.

Table 3.3. Relative Sus outer membrane protein abundances

Protein abundances identified by proteomics were calculated from normalized spectral abundance factor (*NSAF*) values. Total membrane samples 1 and 2 are from *Bt* cells grown on maltose to induce Sus expression. Because of differences in sample concentration, these replicates are not averaged. Co-IP reactions were performed on solubilized fractionated membranes from maltose-grown cells. Co-IP results represent pooled triplicates.

	Total membrane proteome #1	Total membrane proteome #2	SusD Co-IP Native #1	SusD Co-IP Native #2	SusD Co-IP Formaldehyde	SusE Co-IP Native	SusE Co-IP Formaldehyde
SusC	0.80	0.85	0.58	0.60	0.32	NA	0.07
SusD	1.00	1.00	1.00	1.00	1.00	NA	0.21
SusE	0.41	0.52	NA	0.11	0.22	1.00	1.00
SusF	0.41	0.58	NA	NA	NA	NA	NA
SusG	0.23	0.26	NA	NA	NA	NA	NA

In these experiments, the lack of SusG or SusF captured via co-IP was not necessarily unexpected. We have previously suggested that SusG interacts with the Sus complex transiently, and therefore is unlikely to be pulled down with the complex natively [79]. However, here SusG is still not captured with SusD or SusE in formaldehyde-treated cells. One possibility is that formaldehyde does not crosslink SusG because of the absence or spacing of appropriate chemical reactive groups at the interaction interface [122]. Another important distinction is that, for co-IP and proteomics, *sus* expression was induced in *Bt* cells with the small sugar maltose and not on starch as this polysaccharide is difficult to remove from cells after culturing for the downstream experiments. Thus, the entire Sus complex may not associate unless the large starch polysaccharide is present. In addition, SusF may not be captured via co-IP because it is not efficiently solubilized from the membrane, yet it is possible that that SusF in its native environment on the cell surface interacts with the other Sus proteins. How the Sus proteins interact with each other in the membrane is still unknown, though our work demonstrates that SusCDE interact, and our previous work suggests that SusG interacts transiently with these proteins during starch catabolism [79].

Two crystal structures of homologous Sus-like complexes from *Bt* have been determined as reported in Glenwrightd et al. [83]. In the first Sus-like complex, which may target peptides, when the SusCD-like pair was isolated from solubilized membranes, two additional lipoproteins, encoded within the same locus and akin to SusEF, co-purified as part of this complex. The intimate association of these four proteins, BT2261-2264, is one example of Sus-like proteins that may associate stably.